Svoluji k zapůjčení své diplomové práce ke studijním účelům a prosím, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně ocitovat.

Univerzita Karlova v Praze

Přírodovědecká fakulta

Studijní program: Genetika, molekulární biologie a virologie

Studijí obor: Virologie

Bc. Kristian Kříž

Neklasické nekovalentní interakce v proteinech a jejich význam pro návrh nových specifických inhibitorů virových enzymů

Nonclassical noncovalent interactions in proteins and their importance for design of novel specific viral enzyme inhibitors

Diplomová práce

Školitel: RNDr. Martin Lepšík, PhD.

Praha, 2016

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Decleration:

I declare that I carried out this masters work independently and that I cited all of the used sources. This work neither its substantial part has been used to obtain other or the same academic degree.

V Praze dne 28. 4. 2016

Podpis

Poděkování:

Srdečně děkuji svému školiteli, RNDr. Martinu Lepšíkovi, PhD, konzultantovi, RNDr. Jindřichu Fanfrlíkovi, PhD a rovněž každému ze skupiny Prof. Ing. Pavla Hobzy, DrSc., FRSc, D.h.c., za pomoc při zpracování této práce.

Acknowledgement:

I am much obliged to my supervisor Martin Lepšík and the consultant Jindřich Fanfrlík as well as to everyone of Prof. Pavel Hobza DSc. group for being helpful at elaborating this work.

Abstrakt

Nekovalentní interakce jsou zásadní pro fungování biologických systémů. Uplatňují se například při párování bazí DNA nebo při sbalování proteinů. Kromě klasických nekovalentních interakcí jako je např. vodíková vazba jsou v poslední době objevovány i nekovalentní interakce neklasické. Příklad takové interakce může být halogenová vazba, patřící mezi σ-dírové interakce, jejíž znalost se již používá pro optimalizaci a návrh medicínských přípravků. Otázka, kterou se zabývá tato práce, je, zdali také chalkogenová vazba, rovněž σ-dírová interakce, hraje roli při vazbě existujících virových inhibitorů. Dále je také předmětem zájmu, jestli je možné nebo do jaké míry lze tyto existující chalkogenové vazby vylepšit, tj. zvýšit afinitu k virovým proteinům, které váží. Několik komplexů proteinů s ligandy s geometrickým uspořádáním vhodným pro chalkogenovou vazbu bylo nalezeno v PDB (Protein Data Bank) databázi. Pomocí kvantově chemických výpočtů byla zjištěna pro modelové systémy odvozených z těchto krystalových struktur jejich interakční energie a závislost interakční energie na geometrii. Dále jsme sérií substitucí optimalizovali tyto modelové systémy směrem k silnější chalkogenové vazbě. Na základě těchto výsledků považujeme chalkogenovou vazbu za nadějnou neklasickou nekovalentní interakci, která by se dala využívat pro návrh nebo optimalizaci inhibitorů virových enzymů.

Klíčová slova:

Nekovalentní interakce, ligand, inhibitor, virus, chalkogenová vazba, σ-díra, kvantová chemie

Abstract

Noncovalent interactions are vital for functioning of biological systems. For instance, they facilitate DNA base pairing or protein folding. Recently, in addition to classical noncovalent interactions such as hydrogen bond, nonclassical noncovalent interactions have been discovered. An example of these interactions is halogen bond belonging to the class of σ-hole interactions, the knowledge of which is already being useful for medical compound design. The aim of this work is to find out if the chalcogen bond, also a σ-hole interaction, plays a role in the binding of existing viral inhibitors, too. Following that, we are also interested whether or to what extent can these existing chalcogen bonds be optimized for a greater affinity of the inhibitor binding. Several protein-ligand crystal structures exhibiting geometrical properties favoring a chalcogen bond have been found in the PDB database. We examined the interaction energies and the interaction energy geometrical dependencies of model systems derived from these crystal structures by means of quantum chemical calculations. Further we have optimized their strength by a series of substitutions. We thus propose that chalcogen bond can become a player in rational design of inhibitors of viral enzymes and their protein target.

Keywords:

Noncovalent interaction, ligand, inhibitor, virus, chalcogen bond, σ-hole, quantum chemistry

Abbreviations

Å – ångström (the measure of distance, 1 Å=10^{-10} m)

a.u. – atomic unit (the measure of electronic charge, 1 e^-)

BLAST – Basic local alignment search tool (bioinformatic tool)

DFT – density functional theory (the computational method)

ESP – electrostatic potential (the potential of electric field)

FASTA – (amino acid or nucleotide sequence format)

HF – Hartree-Fock (the computational method)

kcal/mol – kilocalorie per mole (the measure of energy, 1 calorie=4.2 Joule)

kJ/mol – kilojoule per mole (the measure of energy)

PDB – protein data bank (the online database of biomolecular structures)

TPSS – Tao, Perdew, Staroverov, and Scuseria functional (the fuctional for the DFT computations)

TZVP – valence triple zeta plus polarization basis set (the basis set for the DFT computations)

Contents

1. Introduction

Formation or breaking of covalent bonds in biological systems is basically irreversible. However, function of biomolecules requires reversible processes. Thus, contrasting to chemistry, which studies formation and breaking of covalent bonds, for biology and related sciences nocovalent bonding is much more important. Indeed, structure of proteins, lipids, sugars or nucleic acids is largely dependent on noncovalent forces. Also fundamental molecular processes such as DNA replication, transcription and translation rely strongly on these interactions.

Noncovalent interactions are about 10-100 fold less stable than covalent bonds. In certain cases, however, rather stronger association of biomolecules is required. For such cases, multiple noncovalent interactions can be combined together to produce a binding so strong that the association of molecules is close to irreversible. Folding of proteins may be an example of multiple noncovalent bonds working together in order to produce a very stable native conformation. For the formation of an α-helix in newly folding protein hydrogen bonds between C=O carbonyl groups and NH amide groups one loop below are essential. Similar to the α-helix, the β-sheet utilizes also carbonyl to amide hydrogen bonds, but this time each moiety is belonging to the different sheet. Further compacting and adoption of a tertiary and quarternary structure of a protein is driven by the formation of noncovalent interactions of protein´s side chains such as salt bridges and hydrogen bonds. The protein structure is then additionally stabilized by van der Waals forces and the hydrophobic effect.

Without a reversible association of both DNA strands, the replication in the form that we know today would not be possible. This is also true for DNA-RNA hybrid in transcription and reverse transcription or RNA-RNA strand in replication of certain viruses. All these associations are determined by specific hydrogen bonds between adjacent bases. Moreover, DNA double helix is stabilized by stacking interaction between aromatic rings of bases placed one over the other.

Aside fundamental processes, indeed, noncovalent interactions are also crucial for protein-protein interactions in allosteric regulation of proteins such as enzymes and for enzyme mediated catalysis itself. Transcription factor-DNA interaction or association of capsid proteins of viruses as well as the formation of a lipid bilayer are of course all governed by noncovalent forces, too.

All the noncovalent interactions mentioned so far are well known and can be considered as classical, well established. The first description of hydrogen bond, at that time in terms of Lewis formalism, dates far back to 1920 (Latimer and Rodebush).

More recently, other kinds of molecular noncovalent interactions have also been described. As for hydrogen bond there are weak hydrogen bonds C-H...O or C-H...π (C-H...aromatic ring), i.e. carbon-attached nonpolar hydrogen with oxygen or aromatic ring interactions. There is also a nonclassical variant of hydrogen bond between partially positively and partially negatively charged hydrogen atoms – the dihydrogen bond (Custelcean and Jackson, 2001). Other example of nonclassical noncovalent interactions are σ-hole based interactions such as halogen or chalcogen bond, that have been described relatively recently (Clark et al., 2007).

The 20th century has witnessed a revolution in the discoveries and development of medical compounds as drugs. The understanding of a disease and a drug action on the molecular level enabled by the progress in the molecular biology lead to the rationalization of the drug development process. The drug binding is mostly based on noncovalent interactions, while the formation of covalent complexes by inhibitors and their targets is rather scarce (Krogsgaard-Larsen et al, 2010). Taking antivirotics as an example, we mention here the development of the anti-influenza drug Tamiflu or, the second generation drug against HIV-1, amprenavir ("Agenerase"). A good understanding of noncovalent forces that govern the drug binding is beneficial for a succesful design of new, optimized compounds.

Alongside with classical, nonclassical noncovalent interactions are currently being investigated as another potential source of affinity gain for ligand binding. The dihydrogen bond for example plays a role in the binding of a certain metallacarborane, another anti-HIV drug (Fanfrlík et al., 2008).

Nonclassical noncovalent σ-hole interactions are today the matter of intense research at both experimental and theoretical level. From these, for example, halogen bond is suggested to improve

the binding affinity of BI 201335, inhibitor of HCV protease NS3, compared to its unsubstituted analog (Lemke et al., 2011). Likewise, halogen addition to HIV-1 reverse transcriptase inhibitors was found to be important for the binding to their target (Himmel et al., 2005).

2. Literature Review

2.1 Physical concepts

To better understand the following sections about nonclassical noncovalent and σ-hole interactions, it is convenient to remember some of the fundamental physical concepts. Introduction of classical noncovalent interactions, hydrogen bond and ionic interactions in section 2.2 serves as a basis for the comparison of their properties to nonclassical noncovalent interactions discussed further in section 2.3. Nonrigorously defined van der Waals forces that included some of forces or effects mentioned in the subsection 2.1.2 will be used in historical context and rather the description of these forces separately will be preferred.

2.1.1 Atomic and molecular properties

2.1.1.1 Charge distribution

Electrostatic properties of molecules are very important for determining the way they interact one with another. In molecules, the electric charge (the magnetic element in binding of two molecules is neglected as unimportant) is rarely distributed evenly, as it is in small spherical monoatomic ions such as Na^+. These ions represent monopoles. In other cases, there are accumulations of positive (low electronegativity) or negative (high electronegativity) charge in certain regions around atoms, electric charge is distributed anisotropically. Electronegativity is the property providing a prediction about the position of such regions of positive or negative partial charge and the magnitude of this partial charge. The charge separation in frame of molecules is described in terms of multipole moments (dipole moment for dipole and so forth). The separation of partially positively and negatively charged regions may results in the electric dipole - such as in a water molecule (Fig. 1). The quadrupole is a separation of two partially positively and two partially negatively charged regions (Fig. 1) (Stone, 2013; Murray and Politzer, 2011).

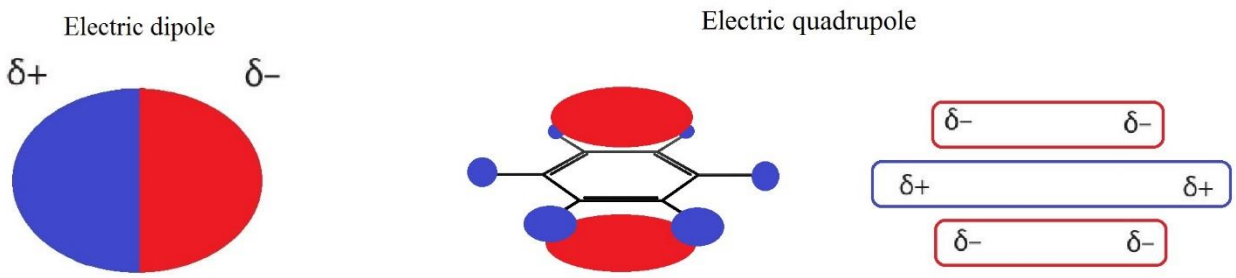


Figure 1. Electric dipole and quadrupole

An example of schematic ellipsoid electric dipole is on the left. Electric quadrupole of a benzene is in the middle and schematically on the right. Partial positive charge δ^+ (blue) is around the hydrogen atoms on the periphery, partial negative charge δ^- (red) is above and under the aromatic ring.

2.1.1.2 Electrostatic potential

The electrostatic potential (ESP) is a potential energy that results from Coulomb forces (Eq. 1, section 2.1.2.1) and it is a physical observable. ESP can be computed for different isosurfaces of a molecule with different shapes and distances from the nucleus. One possible way where to calculate such isosurface is at a certain electron density away from the nucleus. ESP can be measured in units of energy. It has been shown, that ESP can be used for the prediction of the interactional tendencies of certain compounds (Murray and Politzer, 2011).

2.1.1.3 Polarizability

The polarizability describes the ease with which will the electronic cloud of a molecule be influenced by an external electric field and how large can be the dipole that arises from such action (induction). Polarizability is the largest for large molecules with a lot of electrons as it is proportional to the molecule´s radius (or to a distance from the nucleus in the respective axis).

2.1.1.4 Orbitals

According to quantum physics, it is not possible to define the position of a subatomic object with the absolute certainty. Atomic orbitals are areas in space with the greatest probability (arbitrarily defined) that electrons are situated there. Their shape can be derived from the solution of the Schrödinger´s equation as either *s*, *p*, *d* or *f*. Only the *s* orbital is spherically symmetric, as would be intuitively expected. *p* orbital for instance has three lobes along each of axes x, y and z (Figs. 2, 12). When two atoms interact to form a molecule bound by a single bond, their respective orbitals combine into one bonding orbital and one antibonding orbital (two p_z lobes shown in Fig. 2). The

electronic occupancy of the bonding orbital favors the binding while the electronic occupancy of the antibonding orbital weakens the bond.

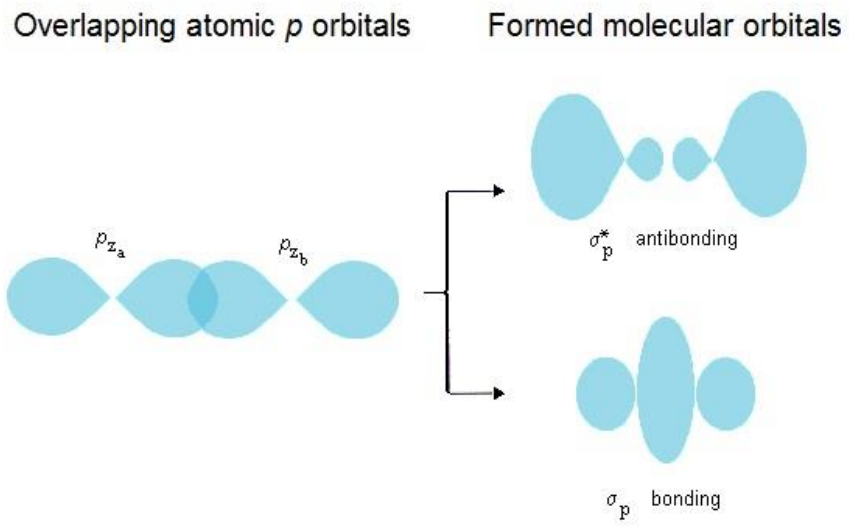


Figure 2. Formation of molecular orbitals by p_z to p_z single bond

In this scheme, two "lobes" of atomic p_z orbitals of compounds (atoms) a and b are brought together to form a single covalent bond. By their combination two different molecular orbitals arise – bonding and antibonding.

2.1.1.5 van der Waals radius

Historically, there was a practical necessity to establish a volume around atoms (molecules) that would not be accessible by other atoms, at ordinary temperatures and pressures, that are not covalently bound to them – it was named van der Waals volume. Such unpenetrable areas were in approximation considered as symmetrical spheres and their radius is van der Waals radius. Moving two atoms together forces their electron clouds to overlap. The Pauli exclusion principle that dictates that only two electrons with the opposing spin can occupy a single orbital (such as p_x) is then manifested in a repulsion (later referred to as exchange repulsion). Thereby the van der Waals radius is a distance between two atoms at which this repulsion balances attractive forces (Fig. 3). Van der Waals radii were obtained from crystallographic data or estimated from covalent bond lengths or other physical properties when necessary (Bondi, 1964). However, as explained later, much of noncovalent interactions also display an intermolecular distance below the sum of van der Waals radii. Van der Waals radii for oxygen, sulfur and selenium are 1.52, 1.8 and 1.9 Å respectively.

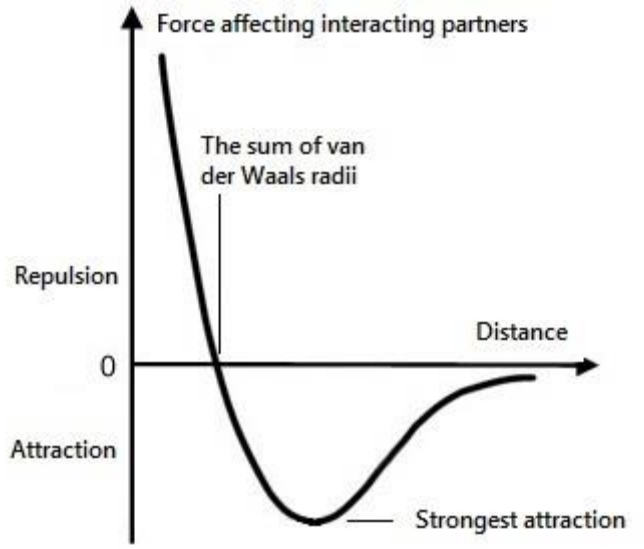


Figure 3. The forces affecting interacting partners as a function of distance

The general function depicting the dependence of forces on the distance of interacting partners can be plotted. Such curve has an energy minimum at a distance where the net attractive forces are the greatest. At the greater intermolecular distance forces affecting the interacting partners are smaller, but always negative - attractive. With the increasing separation forces are asymptotically converging to the x axis (to zero net force). With the decreasing distance, however, repulsive forces start to take over as the net force starts to converge asymptotically to the y axis (to infinite positive interaction energy). The sum of van der Waals radii of two interacting partners is at the point where attractive forces balance repulsive forces.

2.1.2 Physical forces as basis for noncovalent interactions

2.1.2.1 Electrostatic force

Electrostatic force is a force between two (or more) static charge distributions of interacting compounds. Electrostatic force can be either attractive or repulsive (Fig. 4). The force of repulsion or attraction of charged moieties is described by well-known Coulomb equation:

$$|F| = k\frac{|qQ|}{r^2}$$

Equation 1. The electrostatic force

The electrostatic force F is proportional to the product of both charge values of two atomic or molecular species (q, Q) divided by the squared distance (r) between them. The k is the Coulomb´s constant and describes the influence of environment.

From Eq. 1 it is obvious that the force weakens with the second power of the distance between charged compounds, this makes electrostatic force the longest ranged and the strongest relatively to other interactions or forces such as dispersion (Stone, 2013; Kodíček and Karpenko, 2013).

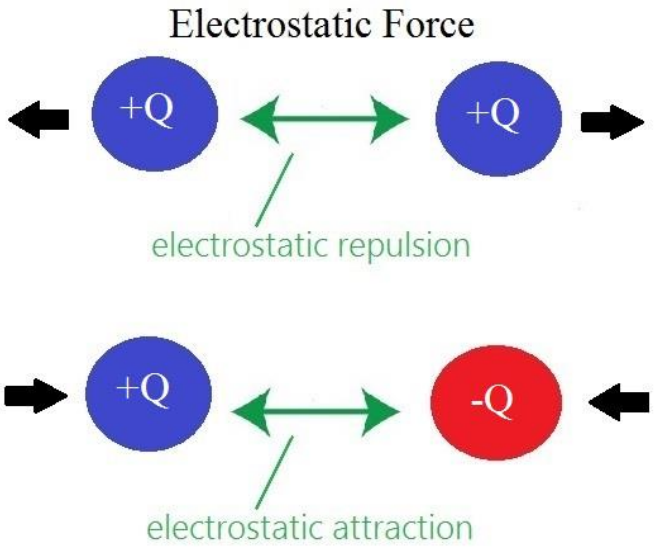


Figure 4. Electrostatic force

Two charged compounds (ions) of the same sign of the charge electrostatically repel each other, opposite signs attract each other.

2.1.2.2 Induction

Induction or polarization is an effect in which a permanent dipole (or a multipole) of one molecule induces a distorsion of electron distribution on the other molecule by means of its electric field and ultimately generates an induced dipole, which can further interact (Fig. 5) (Sedlak et al., 2015; Kodíček and Karpenko, 2013).

2.1.2.3 Dispersion

Dispersion or London dispersion force, named after F. London, is an induced dipole – induced dipole interaction, where the fluctuation of electron cloud of one molecule affects the electron cloud on the other molecule with its electric field. Resulting fluctuations on both molecules are aligned in a way that both molecules attract one another, dispersion is always attractive (Fig. 5, bottom). This interaction or force was historically included in van der Waals forces together with Keesom (dipole-dipole) and Debye (charge-induced dipole or dipole-induced dipole) forces. Nowadays van der Waals interaction and dispersion are often used interchangeably, though

(Grimme, 2011). These interactions wane with the sixth power of distance, and thus are relatively short-ranged. Dispersion is often exemplified by the possibility to turn noble gases liquid (Kodíček and Karpenko, 2013) and is the main stabilizing factor in interactions between neutral and nonpolar moieties. The fluctuation of electronic field is not excluded for polar or charged compound as well, though. Dispersion depends on the polarizability of both molecules, since polarizabilities describe the ease with which induced dipole moments arise due to the electric field (Sedlak et al., 2015).

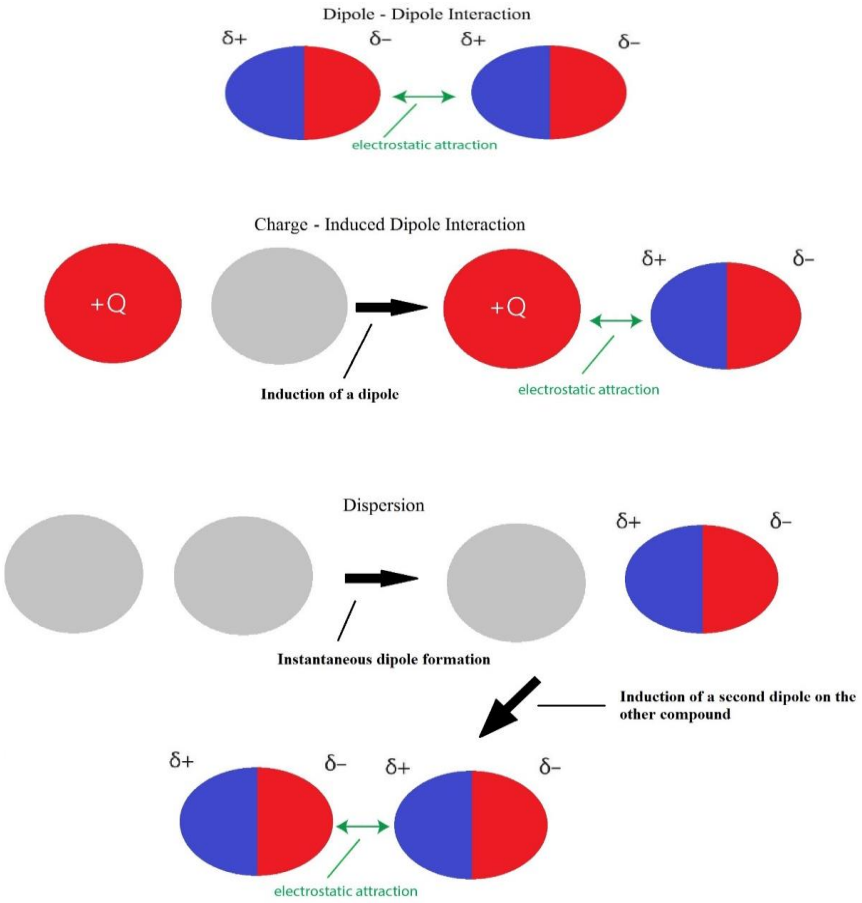


Figure 5. Dipole-dipole, charge-induced dipole and induced dipole-induced dipole interactions

The figure shows schematically three types of interactions historically known as van der Waals forces. Keesom force, dipole-dipole interaction occurs between a positive dipole of one compound with a negative dipole of another. Debye force, charge-induced dipole (or dipole-induced dipole) interaction is an interaction between a dipole induced on one compound by an electric field of the other compound and the other compound. Dispersion is an interaction between an instantaneous dipole and a dipole of an opposite partial charge it induced on the other compound.

2.1.2.4 Charge transfer

Charge transfer is an important effect at close distances where electron clouds of two molecules overlap. It may occur between neutral (nonpolar) compounds A and B and result in $A^{+\delta}$ and $B^{-\delta}$ compounds being attracted by electrostatics. The low ionization energy of A (the ease to become a cation) and the high electron affinity of B (the ease to become an anion) facilitate the transfer (Sedlak et al., 2015; Stone, 2013). Charge transfer is hard to define rigorously and energy decomposition methods define it arbitrarily. Charge transfer is often understood as a covalent component of the binding (Hobza and Müller-Dethlefs, 2010; Palusiak, 2010; Reed et al., 1988; Grabowski et al, 2013a), as in the cases discussed in further sections the charge is transferred to an empty orbital of an acceptor atom and is shared between both atoms. Most noncovalent bonds have a charge transfer component as their electron clouds overlap (Hobza and Müller-Dethlefs, 2010). Charge transfer can facilitate noncovalently interacting atoms to be at the distance below the sum of their van der Waals radii.

Concluding remarks

Nomenclature of these physical forces, effects, is somewhat arbitrary and they are not completely distinct one from another. Electrostatic force, an interaction between charges, is the key player in the other effects mentioned here. This electrostatic force is bound by laws of quantum physics and gives birth to induction, dispersion and charge transfer. These effects together build up more complex phenomena, like hydrogen or halogen bond, that are always referred to as interactions.

2.2 Classical noncovalent interactions

2.2.1 Hydrogen bond

A classical hydrogen bond D-$H^{\delta+}$... $^{\delta-}$A (Fig. 6) is an interaction that is formed between a partially positively charged hydrogen in D-H (where D is an electronegative atom) and a lone pair of an electronegative element A. The D-H group is a hydrogen bond donor and A is a hydrogen bond acceptor. In Lewis acid-base formalism, A is an electron donor (Lewis base) and D-H is an electron acceptor (Lewis acid). Less conventional, weak hydrogen bond can also be formed with a weakly

polar carbon attached hydrogen atom as a hydrogen bond donor (C-H…O) or with a π-aromatic system as an acceptor (C-H…π) (Sedlak et al., 2015). From different perspective, hydrogen bonds can be classified either as neutral (as in a water dimer), charge assisted (carbonyl oxygen to NH_3^+) or ionic - "salt bridge" (both interacting partners are charged) (Krogsgaard-Larsen et al, 2010). Electrostatics provides a substantial amount of stabilization in hydrogen bond, but hydrogen bond has also a covalent component (Hobza and Müller-Dethlefs, 2010; Reed et al., 1988). Significant part of interaction energy comes from charge transfer as the electronic charge is transferred from the electron donor to the antibonding σ of the respective D-H bond. Hydrogen bond is thus considered as partly covalent. This charge transfer is accompanied by D-H bond lengthening (Hobza and Müller-Dethlefs, 2010; Reed et al., 1988). The partial covalency of hydrogen bond is also supported by the fact, that hydrogen bond distances can be below the sum of van der Waals radii as a typical hydrogen bond length (D...A distance) is between 2.5 and 3 Å (Krogsgaard-Larsen et al., 2010). The hydrogen bond D-H...A is most stable as linear – with the D-H...A angle of 180° (Bissantz et al., 2010; Sedlak et al., 2015). The interaction energy of a typical hydrogen bond lies between -3 and -8 kcal/mol (Kodíček and Karpenko, 2013).

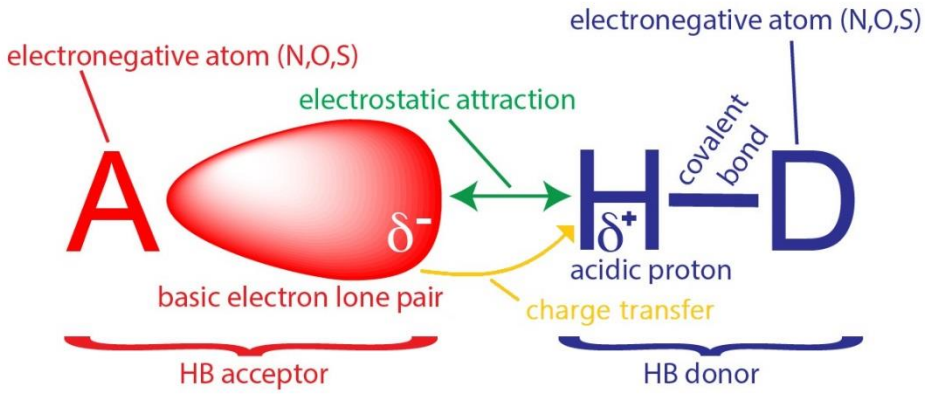


Figure 6. Schematic view of hydrogen bond

"A" denotes an electronegative atom, a hydrogen bond acceptor – electron donor. D-H group is a hydrogen bond donor – electron acceptor, where "H" is a hydrogen atom and "D" is a general electronegative element. Electrostatic interaction between partially positively charged, acidic, hydrogen atom and partially negatively charged, basic, A, or its lone electron pair is in green. Charge transfer from lone pair of electronegative A to the H is in yellow. Modified from Loren Williams, http://ww2.chemistry.gatech.edu/~lw26/structure/molecular_interactions/mol_int.html.

2.2.2 Ionic interaction

Ionic interaction results from the electrostatic force between two permanently charged compounds, ions - for instance between Mg^{2+} ion binding to an aspartate side chains in a polymerase active site. However in a polar solvent such as water the ionic interaction is dampened due to the higher environment permitivity (Kodíček and Karpenko, 2013), 80 times larger than in the vacuum (http://ww2.chemistry.gatech.edu/~lw26/structure/molecular_interactions/mol_int.html).
Similarly, the energy of an interaction between two ionic molecules pair suffer considerable energy loss because of the desolvation penalty (see section 3.3.3.6.5) (http://www.cambridgemedchemconsulting.com/resources/solvation.html).

2.3 Nonclassical noncovalent interactions

2.3.1 Dihydrogen bond

Dihydrogen bond is unconventional D-$H^{\delta+}$...$^{-\delta}$H-B interaction where D is a general electronegative hydrogen bond donor and B is an element even less electronegative than H, further on boron is considered (Custelcean and Jackson, 2001). Despite being considered less conventional and is generally less known, the occurrence of interaction of such partially negatively and positively charged hydrogen atoms was first noted by Zachariasen and Mooney early in 1934 (Zachariasen and Mooney, 1934). The stabilization energy of a typical dihydrogen bond of about 6 kcal/mol and the $H^{\delta+}$...$^{-\delta}$H bond length (1.7-2.2 Å) are both comparable to the classical hydrogen bond (Richardson et al., 1995). Dihydrogen bond has mostly electrostatic character with the dispersion component (Sedlak et al., 2015). The possibility of charge transfer from the bonding σ orbital of $^{\delta-}$H-B to the antibonding σ orbital of D-$H^{+\delta}$ has been proposed (Grabowski, 2013b, Zierkiewicz and Hobza, 2004). This is consistent with the observed directional tendencies, where the D-H vector was pointing more to the H-B bond rather than to the hydrogen atom bound to the boron atom directly (Richardson et al., 1995; Cramer and Gladfelter, 1997). This fact together with distances well below the sum of van der Waals radii in dihydrogen bonds support partial orbital overlap and charge transfer. Lengthening of the D-H bond interacting with H-B was observed (Cramer and Gladfelter, 1997) and analytically examined (Zierkiewicz and Hobza, 2004), suggesting the electron transfer to the antibonding orbital of H-B. Strong dihydrogen bond was found in the

complex between a boron cluster compound and a human carbonic anhydrase II (Fig. 7) (Pecina et al., 2013).

Figure 7. Dihydrogen bond

The boron cluster interacting with an asparagine side chain in the active site of human carbonic anhydrase (Adopted from Pecina et al. (2013). Interatomic distance between asparagine´s oxygen atom and boron cluster´s hydrogen atom is in Å. Boron atoms are pink, carbon green, hydrogen white, nitrogen blue, sulfur yellow and oxygen atoms are red.

2.3.2 Aryl-aryl interaction

Aryl-aryl, aromatic-aromatic or π-π stacking interaction is a common theme in proteins and nucleic acids as these polymers contain many aromatic rings (bases of nucleic acids and side chains of Phe, Tyr, His and Trp amino acids). Aromatic rings have large quadrupole moments permitting significant electrostatic interaction. Additionally, dispersion plays an important role for these types of interactions. Aromatic benzene has a partial negative charge above and below the ring and a partial positive charge on the periphery (see Fig. 1 middle and right). Such properties favor off-center parallel stacking (or displaced parallel stacking), where the center of one aromatic ring is shifted from the above of the center of another stacked ring, such as in DNA double helix (Fig. 8, top). Another preferred motif is edge-to-face stacking or interaction, where the ring hydrogen atom interacts with the electron cloud above the other ring (Fig. 8, bottom). The latter belongs to the weak C-H…π hydrogen bond mentioned above (but sometimes is considered as a stacking interaction, too). For aromatics substituted with electronegative moieties (fluorine atoms or cyano groups), the quadrupole moment may be inverted leaving the partial negative charge on the periphery and the positive charge in the center. This facilitates the molecule to interact more favorably with a typical electron rich-centered aromatic ring in face-to-face (center above center) manner (Martinez and Iverson, 2012; Bissantz et al., 2010).

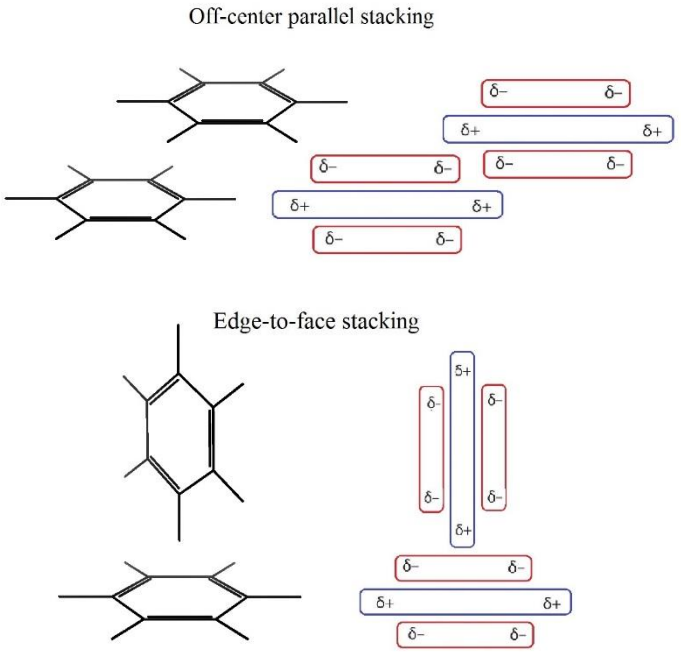


Figure 8. Interactions of two benzene rings

For unsubstituted benzene, energetically the most favorable orientations of two aromatic rings are off-centered parallel stacking (top) and edge-to-face stacking (bottom). The former is a π-π interaction, the latter belongs to the class of hydrogen bonds.

2.3.3 σ-hole interactions

Among σ-hole interactions the largest amount of work has been done on the halogen bond and therefore it will be the main topic of this section. The less widely studied chalcogen bond is the main topic of this work and as such will be also discussed here. Other σ-hole interactions, such as pnictogen and tetrel bond interactions, are not the subject of this work and will be mentioned only briefly.

2.3.3.1 Historical perspectives

Complexes of heavier VII group elements (Cl, Br, I), and amines have long been known. For instance, Remsen and Norris observed the formation of $(CH_3)_3N...Br_2$ crystals in 1896 (Remsen and Norris, cited by Bent, 1968). In 1954 Hassel and Hvoslef ´s crystallographical study of 1,4 dioxane and Br_2 complexes revealed peculiarly short interatomic distance between bromine and oxygen atoms in their respective molecules - 2.71 Å in the crystal contrasting with 3.35 Å that is the sum of van der Waals radii for respective elements (Fig. 9) (Hassel and Hvoslef, 1954, cited by

Bent, 1968). Bromine atom was apparently participating in the Br...O interaction. The axis of the Br-Br bond was facing the oxygen atom – in other words, the Br-Br...O angle was close to 180°.

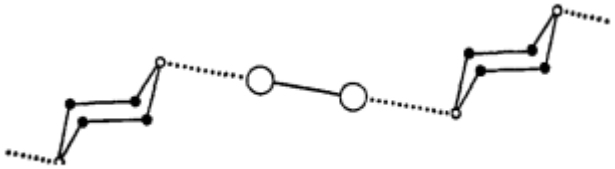


Figure 9. Complex of Br_2 and 1,4-dioxane

Filled circles – carbon atoms; small empty circles – oxygen atoms; big empty circles – bromine atoms. An orientation of a bromine molecule so that the Br-Br axis is facing an oxygen atom is apparent. Adopted from Bent et al. (1968).

Many more structures containing this common theme have been observed since (Damm et al., 1966, cited by Bent, 1968; Gagnaux and Susz, 1960, cited by Bent, 1968; Groth and Hassel, 1963, cited by Bent, 1968; Murray-Rust and Motherwell, 1979). Similarly, sulfur (group VI element) to an oxygen atom close contacts were revealed to be relatively abundant in the Cambridge Crystallographic Database (Kucsman and Kapovits, 1985, cited by Burling and Goldstein 1992). Systematic research of this database regarding environment of C-X bonds (X is a heavy halogen atom) revealed a tendency of nucleophiles to approach halogen atoms of C-X by head-on fashion. Electrophiles approached such halogens rather from sides (Ramasubbu et al., 1985). The same tendency was observed in the crystal structure of I_2 (Fig. 10) (Harris et al., 1928). An iodine atom in the crystal was acting as an electron donor towards another iodine atom (of a different I_2 molecule) in 90° of the I-I bond and as electron acceptor in 180°of the I-I bond. It was also shown that such interactions strengthen going from lighter to heavier halogens (Hassel, 1970).

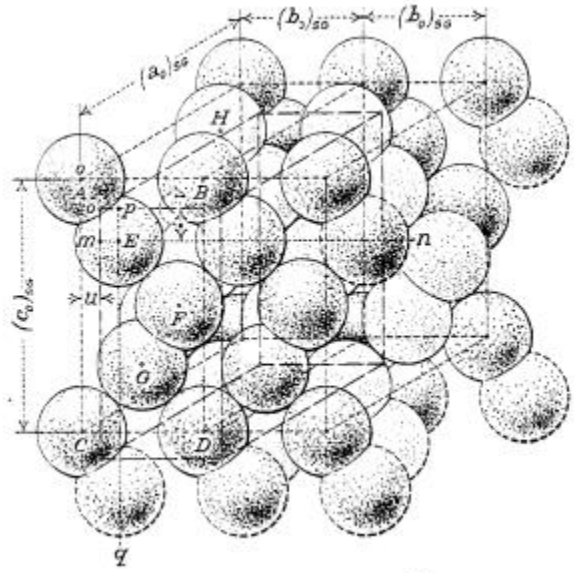


Figure 10. Crystal structure of molecular iodine - I_2

In the crystal of molecular iodine certain nonrandom geometrical orientation is apparently preferred. This suggests that some highly directional interactions are at play. Adopted from Harris et al. (1928).

2.3.3.2 Modern interpretation

Halogen atoms interactions with electrophiles (positively or partially positively charged moieties) were not unexpected since halogen atoms belong to the most electronegative species in the whole periodic table. However the idea that halogens (as parts of their respective molecules) could interact with nucleophiles (negatively or partially negatively charged molecules) was rather unintuitive. In 1992, regions of positive electrostatic potential on otherwise negatively partially charged bonded halogen containing compounds were discovered computationally (Brinck et al., 1992). This phenomenon of halogens being able to interact with a nucleophile is today referred to as halogen bond, as it was named by Dumas et al. (Dumas et al., 1976, cited by Politzer et al., 2013).

The interpretation of such interactions was at first revolving around the charge transfer concept, where the negative charge was to be transferred from an electron donor to a halogen atom (Mulliken and Person, 1969; Hassel, 1970). In case of sulfur-oxygen close contacts (chalcogen bond, see below), interaction of oxygen lone pair with the antibonding S-Y (S is sulfur) σ orbital along with the electrostatic attraction was proposed quite early. The electrostatic attraction was provided by the chalcogen atom - partially positively charged due to the influence of the rest of the

molecule (Becker et al., 1986; Burling and Goldstein et al., 1992). These ideas were close to how the phenomenon is perceived today – at least by the majority of researchers.

Later on additionally to positive regions on *bound* halogenated molecules, the analysis of electrostatic potentials of *isolated* halogen containing compounds also revealed positive areas (Fig. 11) (Clark et al., 2007). From there, charge transfer or induction effects caused by interacting partners alone could not explain this phenomenon. This positive region was named σ-hole (see below).

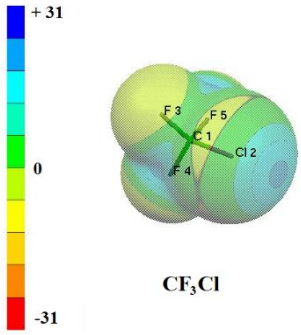


Figure 11. ESP of CF_3Cl

Electrostatic potential of CF_3Cl at 0.001 electrons per $bohr^3$. The color scale from +31 to -31 is in kcal/mol. Except for the σ-hole on the chlorine atom, σ-holes on carbon atoms against the C-F bonds are apparent as well.

The modern interpretation of halogen bond was presented by Clark et al. (2007). They elegantly explained the origin of the charge anisotropy on bonded halogen atoms in terms of quantum physics. In the case of heavier halogen atom where the sp^3 hybridization is not so significant, the *p* valence shell contains 5 electrons distributed by 2 pairs in p_x, p_y orbitals each and one single electron delocalized in the two lobes of the p_z orbital (Fig. 12).

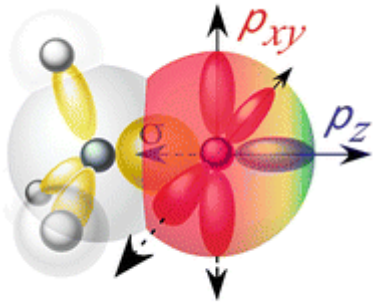


Figure 12. σ-hole along the p_z axis

Schematic electrostatic potential on the halogen atom (right). Green color codes posistive regions, red color codes negative regions. Colocalization of positive ESP and the nonbonding lobe of the p_z orbital is apparent. Adopted from Ho (2014).

Upon binding of X to C (X is a halogen atom, C is a carbon atom) with one p_z electron of X, the single p_z electron is participating in the newly formed C-X σ bond and thereby its occupancy in the other lobe of the p_z orbital decreases (e.g. in the other lobe of the bonding σ_p – see sections 2.1.1.4 and 2.3.3.6.7). This corresponds to the increased partial positive charge on the other side of the halogen atom, opposite to the C-X bond (Figs. 11, 2). Such areas of positive electrostatic potential on overall negatively charged halogen were called σ-holes, since they exist due to the formation of σ bonds on the other side of the halogen (Clark et al., 2007). Indeed, Burling and coworkers were close to this interpretation when they claimed that "This partial positive charge results from donation of p_z electrons to the π-system..."about chalcogen-oxygen close contacts (1992).

The concept of σ-hole was later expanded to chalcogens - group VI atoms (S, Se, Te) (Murray et al., 2007), corresponding to earlier observations of sulfur (selenium) – oxygen complexes, where the chalcogen acts as a σ-hole donating moiety (Burling and Goldstein, 1992). The interaction of a chalcogen with a nucleophile mediated by its σ-hole is known as the chalcogen bond.

Recently, the σ-hole concept has also been established for pnictogens (pnictogen bond), group V atoms (N, P, As, Sb, Bi), and even tetrels (tetrel bond) - group IV elements (C, Si, Ge, Sn, Pb) are in some cases capable of exhibiting a σ-hole (Politzer et al., 2008). As in the previous cases, such interactions of groups IV, V with nucleophiles were already empirically observed before (Carré et al., 1995; Mitzel et al., 1997). Surprisingly, aerogens or noble gases (He, Ne, Ar, Kr, Xe, Rn)

containing compounds might also exhibit σ-hole (Bauzá and Frontera, 2015). Due to the low reactivity of noble gases, such compounds are rarely seen, though.

The theoretical explanation of chalcogen, pnictogen, tetrel or aerogen bonds is equivalent to that of halogen bonds except that these elements have more σ-holes; they may have as many σ-holes as they have covalent bonds (Murray et al., 2008a). In the case of divalent chalcogens, two chalcogen bonds can be formed utilizing two-holes of two corresponding σ binding orbitals.

2.3.3.3 σ-hole interactions in biological systems

The occurrence of halogenated molecules in biological systems is not very frequent, aside from medicals and their metabolites (see section 2.3.3.). Natural halogenated compounds for example are thyroid hormones. Multiple short contacts between an iodine atom of thyroid hormones and an oxygen atom of their protein interaction partner have been observed (Cody and Murray-Rust 1984). Further, halogenated compounds can be created by the action of the immune response against pathogens. Halogenation of bases (for example uridine to bromouridine) can influence the DNA conformation due to the O...Br interactions (Auffinger et al., 2004).

On the other hand, chalcogens do occur in biological systems as sulfur or selenium atoms in proteins (in cysteine, methionine and selenocysteine amino acids). It has been shown that chalcogen bond may stabilize the protein structure, where a divalent sulfur of methionine or cysteine interacts with a carbonyl oxygen (Iwaoka et al., 2002).

Concerning thyroid hormones, it has been proposed, that not only halogen bond, but also chalcogen bond plays a role in deiodation of thyroid hormones, since iodine removing enzymes do contain a selenocysteine in the catalytic center (Manna and Mugesh, 2012). Selenocysteine is also contained within a radical disposing enzyme glutathione peroxidase. The transition state of the reaction facilitated by this enzyme may be stabilized by a σ-hole interaction between the selenium atom in the active site and an oxygen atom of a substrate radical (Zielinski et al., 2014). Chalcogen bond has also been implied to play a role in the reaction mechanism of methyltranferases (Fick et al., 2016). Another example are phospholipases PLA2, where multiple S...O and S...N interactions were suggested, possibly influencing the activity of these enzymes (Iwaoka and Isozumi, 2006).

Even the breakage of disulfide bridges by a nucleophilic attack could also be somewhat influenced by a σ-hole, as results shown by Mó et al. indirectly imply (2013).

2.3.3.4 σ-hole interactions in protein inhibitors

About one third of medically used compounds contain at least one halogen (Wermuth, 2008, cited by Erdélyi, 2012) as halogens were originally incorporated to increase the membrane permeability and the stability of compounds (Parisini et al., 2011). Heavy halides that should have greater σ-holes (see section 2.3.3.6 σ-hole properties) are more scarce, though (Wilcken et al., 2012). It is still reasonable to expect halogen bond to play a role protein inhibitor binding, though. There are several examples of halogen bonds in medical compounds. The selectivity of human aldose reductase inhibitor (for treating complications associated with diabetes) is explained by the compound´s aromatic ring bromine atom interacting with an oxygen atom of a threonine side chain (Howard et al., 2004). One of the first drugs, with an engineered halogen bond, was factor Xa inhibitor – anticoagulans (Lam et al., 2009, cited by Parisini et al., 2011). Bromine to aspartate halogen bond was found to participate in the binding of HCV inhibitor BI 201335 where bromine substitution corresponds to 3.6 fold increment in the binding affinity over an unsubstituted analog (Lemke et al., 2011). Non-nucleoside inhibitors of HIV reverse transcriptase were also improved by halogen substitutions. Addition of iodine atoms to specific positions of R165481, R221239 (Fig. 13) (Himmel et al., 2005) and similar compounds (Benjahad et al., 2003) were probably responsible for a slight improvement of binding properties of these compounds.

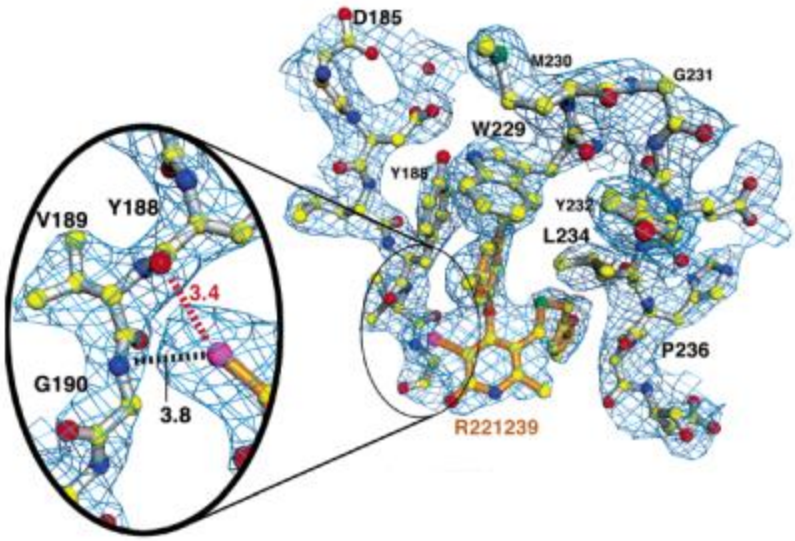


Figure 13. Detail of R221239 binding to HIV reverse transcriptase

Non-nucleoside inhibitor of reverse transcriptase R221239 interacting with the tyrosine 188 (and also the glycine 190) by a halogen bond via its iodine atom. Adopted from Himmel et al. (2005).

Halogen bonding has also been implicated for example in inhibitors of human cathepsin, MEK1 kinase (Hardegger et al., 2011), positive transcription elongation factor b - P-TEFb (Baumli et al., 2010) or p53 regulator MDM2 (Parks et al., 2005; Grasberger et al., 2005).

Studies analyzing chalcogen bond are scarce by themselves and more so for medically relevant systems. Intramolecular sulfur or selenium to oxygen atom close contacts have been implicated to stabilize the conformation of chalcogen containing thiazole and selenazole nucleoside analogs with an antitumor activity (Burling and Goldstein, 1992), increasing their affinity for inosine monophosphate dehydrogenase (Goldstein et al., 1985). However, up to date intermolecular interactions between chalcogens and nucleophiles in medically relevant compounds have not been thoroughly examined.

2.3.3.5 Description

Kolář and coworkers provided a simple nomenclature for the description of σ-holes (2014). They defined the *size* of the σ-hole as a (maximum) angle between two points on the boundary of the atom displaying the σ-hole, where the value of ESP (electrostatic potential) changes its sign, and a third point that corresponds to a nucleus of the atom (Fig. 14).

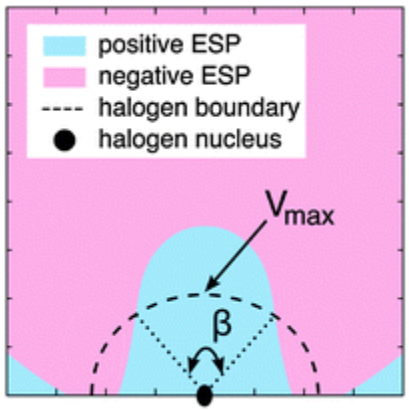


Figure 14. Size and magnitude of a σ-hole

Schematic representation of a σ-hole around a halogen nucleus. Its size is defined by the β angle between the points on the atomic boundary, where the charge sign changes from negative to positive, and a nucleus (the most obtuse angle – so that the plane of this angle is aligned with the p_z axis). The magnitude is the maximum value of ESP in this area. Adopted from Kolář et al. (2014).

The atom boundary corresponds to the electron density 0.001 *e* per bohr^3 (defined earlier by Bader et al., 1987) Thus the σ-hole size tells simply how wide the σ-hole is at certain distance from the nucleus. Kolář and coworkers also introduced the σ-hole *magnitude*, which is the maximum value of ESP within that boundary (V_{max} in Fig. 14). The σ-hole magnitude correlates well with the strength of halogen bonds (Riley et al., 2011).

2.3.3.6 Properties

The halogen bond is the most extensively studied σ-hole interaction. It is presumed that other σ-hole interactions do not differ in critical aspects (Wang et al., 2009). However it is important to keep in mind, that even among halogen bonds there are differences based on the theoretical approach - model system choice and computational setup.

2.3.3.6.1 Strength

The strength of a halogen bond (expressed by the interaction energy) can vary from up to -3 kcal/mol with a suitable orientation in biological systems (Table 1) (Lu et al., 2009) to over -5 kcal/mol in artificial model systems, rivaling the strength of a typical hydrogen bond (Riley and Hobza, 2013).

Table. 1 Halogen bonds in biological systems

Halogen bonds between a protein moiety and a ligand (inhibitor). X is a halogen atom. Electron donors (El. donor) are carbonyl backbone oxygen atoms except for an aspartate 86, that is an oxygen atom of a side chain. Except for the interaction energies, all the values were empirically observed. Systems chosen from Lu et al. (2009). JNK3 - c-Jun terminal kinase 3, CDK2 - cyclin-dependent kinase 2, CK2 - cyclin kinase 2, MEK1, mitogen-activated protein kinase kinase 1, GRP94 - glucose-regulated protein 94, hTRβ - human thyroid hormone receptor.

Note, that systems which deviate too much from the optimal 180° C-X…O angle have positive (i.e. repulsive) interaction energies. Therefore if there was not the rest of a protein, these residues under such angle would probably not be close one to another – no halogen bonding probably occurs in these systems.

PDB ID	Protein	Resolution (Å)	El. donor	Distance X…O (Å)		C-X…O	X…O-C	ΔE (kcal/mol)
1PMN	JNK3	2.20	Ala91 O	2.83	Cl…O	156	96	+2.75
1HIR	CDK2	2.00	Asp86 O	2.84	Cl…O	140	110	+0.34
1ZOH	CK2	1.81	Val116 O	2.89	Br…O	177	129	-2.13
1S9J	MEK1	2.40	Val127 O	3.13	I…O	174	129	-1.60
1QYE	GRP94	2.10	Gly153 O	2.93	Cl…O	142	101	+089
1NQ2	hTR*β*	2.40	Phe272 O	3.07	I…O	171	120	-1.66

Depending on the molecular environment of the interacting residues, halogen, chalcogen bond or hydrogen bond can be the strongest (Fang et al., 2015). In certain model systems, chalcogen bond can have a comparable strength to the halogen bond (Pavan et al., 2015, Zhao, 2014). In other model systems halogen and chalcogen bonds are of comparable magnitudes in terms of interaction energy with the halogen bond being slightly stronger (see Table 2). This result however is electron donor dependent. For instance, when a σ-hole containing compound interacts with trimethylamine the halogen bond is the stronger of the two interactions and when the interacting partner is an aromatic ring, it is the chalcogen bond that is the stronger (Bauzá et al., 2013).

Table 2. Interaction energies of halogen and chalcogen complexes

Interaction energies of chosen systems from Bauzá et al., (2013). Energies are in kcal/mol. Halogen bond in complexes 1, 2, 5 and 6; chalcogen bond in 3, 4, 7 and 8 (see Fig. 15). TMA stands for trimethylamine.

Complex	Energy
1 trifluoromethylbromine - TMA	-7.8
2 trifluoromethyliodine - TMA	-11.2
3 perfluordimethylthioether - TMA	-4.0
4 perfluordimethylselenoether - TMA	-5.9
5 trifluoromethylbromine - benzene	-3.5
6 trifluoromethyliodine - benzene	-4.7
7 perfluordimethylthioether – benzene	-4.8
8 perfluordimethylselenoether - benzene	-5.7

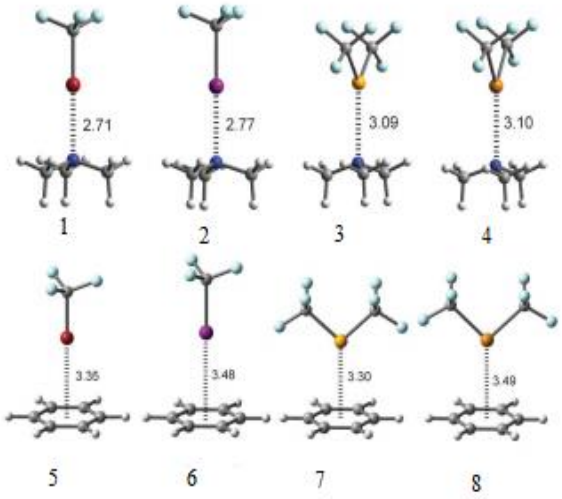


Figure 15. Complexes with halogen and chalcogen bonds

Chosen model systems from Bauzá et al. (2013). White spheres denote hydrogen atoms, gray carbon, light blue fluorine, red bromine, magenta iodine, yellow sulfur and orange selenium atoms.

It is an interesting fact that halogen and chalcogen bonds may be cooperative (Zhao, 2014, Manna and Mugesh, 2012). For example, σ-hole of a halogen binding to an electronegative chalcogen site

can facilitate a formation of another σ-hole (by removal of the electronic charge that is then donated to a halogen σ-hole) on the chalcogen facing the opposite direction (Manna and Mugesh, 2012). Another important thing to consider is that halogen or chalcogen bond and hydrogen bond can be competitive (Syzgantseva et al., 2013) so that partially positively charged σ-hole competes with partially positively charged hydrogen atom for binding of (partially) negatively charged moiety.

2.3.3.6.2 Tunability

The strength of halogen bond depends on adjacent substituent(s) (Fig. 16). The magnitude of σ-hole can thus be "tuned". For example, addition of electronegative fluorine(s) can have a substantial effect on the σ-hole size (Fig. 17) (Riley et al., 2011). With the addition of electronegative substituents, the dispersion component becomes smaller as σ-hole grows and electrostatics predominates (Riley and Hobza, 2008). Concerning substitutions, their position relative to electronegative Y (for instance when Y is not an aliphatic carbon atom, but a part of an aromatic ring) is an important determinant of σ-hole, too (Kolář et al., 2014). Easy manipulation with the σ-hole strength is a factor beneficial for the medical compound optimization.

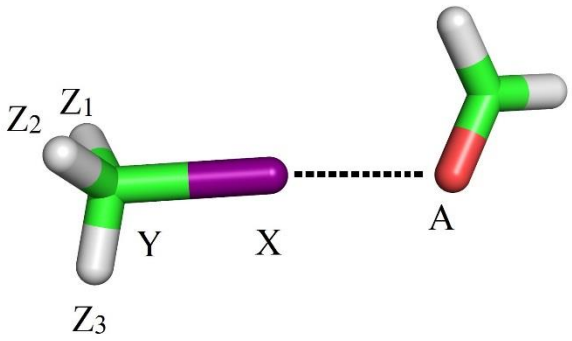


Figure 16. A schematic picture of a halogen bond

Halogen X in X-Y-Z_n molecule interacting with a nucleophile A. In this case an iodine atom of methyliode interacts with an oxygen atom of formaldehyde as a nucleophile. Carbon green, hydrogen white, oxygen red and iodine purple.

It is a general rule, that σ-hole becomes bigger with the increasing atomic number (i.e. going down the column of the periodic table) of the σ-hole possesing atom X (Fig. 16). This was attributed to the growing polarizability (facilitating greater dispersion) and to the decrease in electronegativity (facilitating more positive σ-hole) of σ-hole possessing X (Murray et al., 2014; Riley and Hobza, 2008).

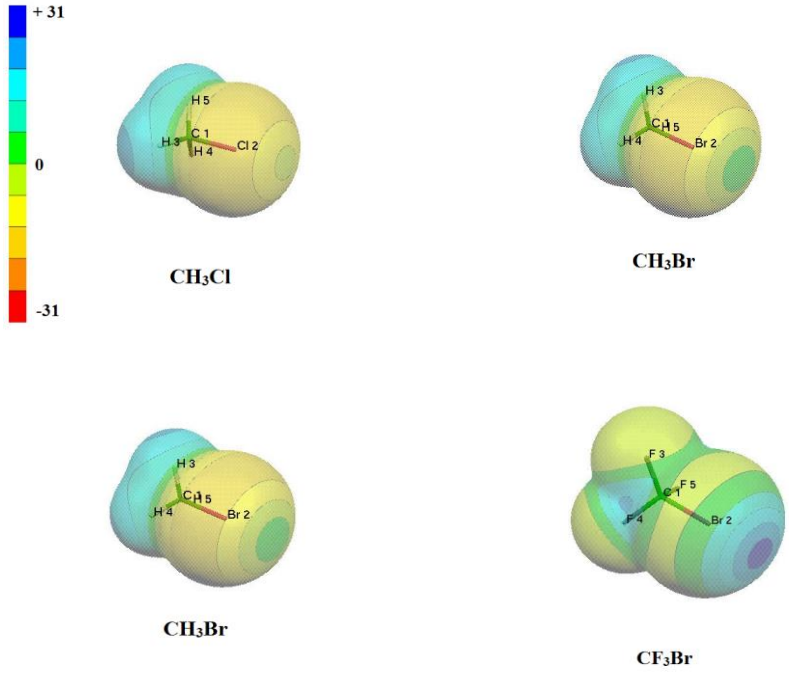


Figure 17. σ-hole tuning, electrostatic potentials of CH_3Cl, CH_3Br, CF_3Cl and CF_3Br.

It is apparent that σ-holes are greater on heavier elements (bromine compared to chlorine atoms) and on molecule substituted with electronegative fluorine atoms (compared to molecules substituted with hydrogen atoms) (Z_n ligands, Fig. 16.). Color scale is in kcal/mol for all the models: blue shades are inclining towards partially positive regions, yellow (red) shades towards relatively more electronegative ones.

Certain authors argued, that the electronegativity of an atom Y alone (Fig. 16) is not a sufficient parameter to describe the X´s capability to generate a σ-hole (Murray et al., 2014). Therefore they implemented a charge capacity (proposed originaly by Huheey, 1965) of Y as a measure to better acomodate the electron gain of Y (drawing electron density from X and increasing the σ-hole respectively). The charge capacity increases with the Y ´s polarizability, which in turn increases with the atomic volume (Politzer et al., 2008). As mentioned above, if Z_n substituents are relatively more electronegative, then the σ-hole on X will be relatively bigger (Murray et al., 2014; Riley and Hobza, 2008). The electron "attractability" again is understood in terms of high electronegativity of Z_n accompanied by high polarazibility of an atom Y. Thus in such system, Y facilitates the electron shift from X to electronegative Z (Murray et al., 2014).

2.3.3.6.3 Distance between interacting partners

The distance of A…X (A is an electron donor, Fig. 16) in σ-hole interactions (e.g. halogen bond) are typically below the sum of van der Waals radii of A and X (Lu et al., 2009) This can be understood in terms of electron deficiency in the nonbonding lobe of p_z orbital so that the missing electron does not prevent further approaching by forces described above (see section 2.1). Interestingly, the σ-hole interaction by heavier atom X can permit closer A…X distance than a lighter element, even though the van der Waals radius of heavier elements is larger (Figs. 18, 19) (Wilcken et al., 2013). The same applies for X substituted with electron withdrawing groups, where the interaction distance is smaller in its optimum compared to unsubstituted X (Riley and Hobza, 2013).

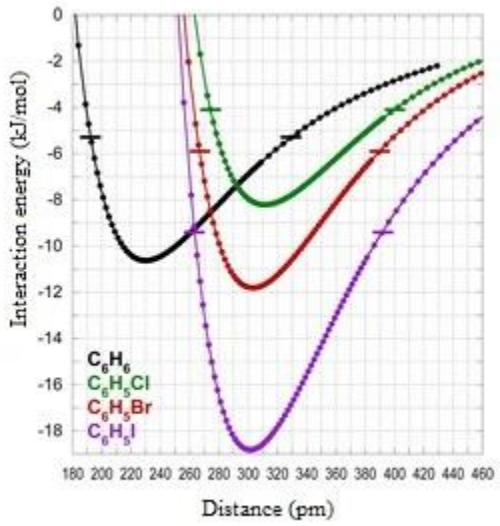


Figure 18. Dependence of the halogen and the hydrogen bond strength on the molecular separation

The distance of carbonyl oxygen (not shown) to hydrogen (black line) or halogen atom (green line chlorine, red bromine and magenta iodine atom) substituted on benzene is on the x axis, in pm. Interaction energies on the y axis, in kJ/mol. Adopted from Wilcken et al. (2013).

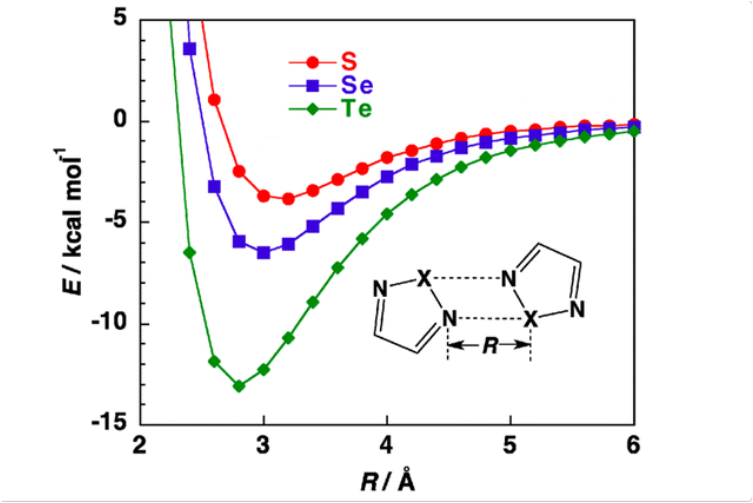


Figure 19. Dependence of the chalcogen bond strength on the molecular separation

The interaction energy (E) in kcal/mol is plotted against the distance (R) between interacting moieties, in ångströms. The strength of the interaction is greater when the bigger chalcogen (X) is involved. Surprisingly, the larger chalcogen exhibits shorter interatomic distances due to large σ-hole. Adopted from Tsuzuki and Sato (2013).

2.3.3.6.4 Directionality

As the σ-hole concept implies, there is an angular preference for the nucleophile A approach to the σ-hole on X (Fig. 16). The A…X-Y angle is close to linear in the energy optimum (Figs. 20, 21) (Wang et al., 2004). The angle distribution for most halogen bonds in biological systems is however broader and further from linearity, between 170-140° (Table 1) (Auffinger et al., 2004; Lu et al., 2009). The deviations from the ideal geometry have been vaguely attributed to the complexity of biological systems (Auffinger et al., 2004).

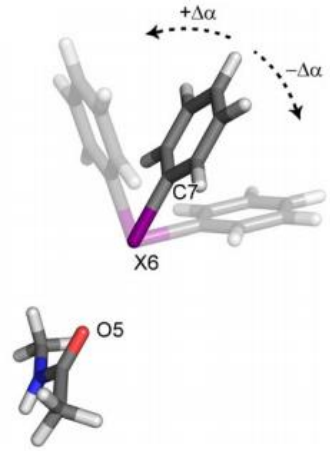


Figure 20. Examining halogen bond complexes with the angular change

Halobenzene interacting with a carbonyl group. The deviation from 180° of C-X…O is designated +Δα and –Δα. Adopted from Wilcken et al. (2013).

For bromine, the interaction energy depends on an A…Br-Y angle. The interaction energy changes gradually for the smaller angles, but then drops more rapidly (stronger interaction) when the angle is above 150° (Fig. 21). Halogen bonds thus should be more directional comparing to hydrogen bonds, in which the relation between the angle and the interaction energy changes rather in a linear

fashion (Shields et al., 2010). The deviation from the ideal geometry by 30° dampen the (model) interaction to about one half (Fig. 21) (Wilcken et al., 2013).

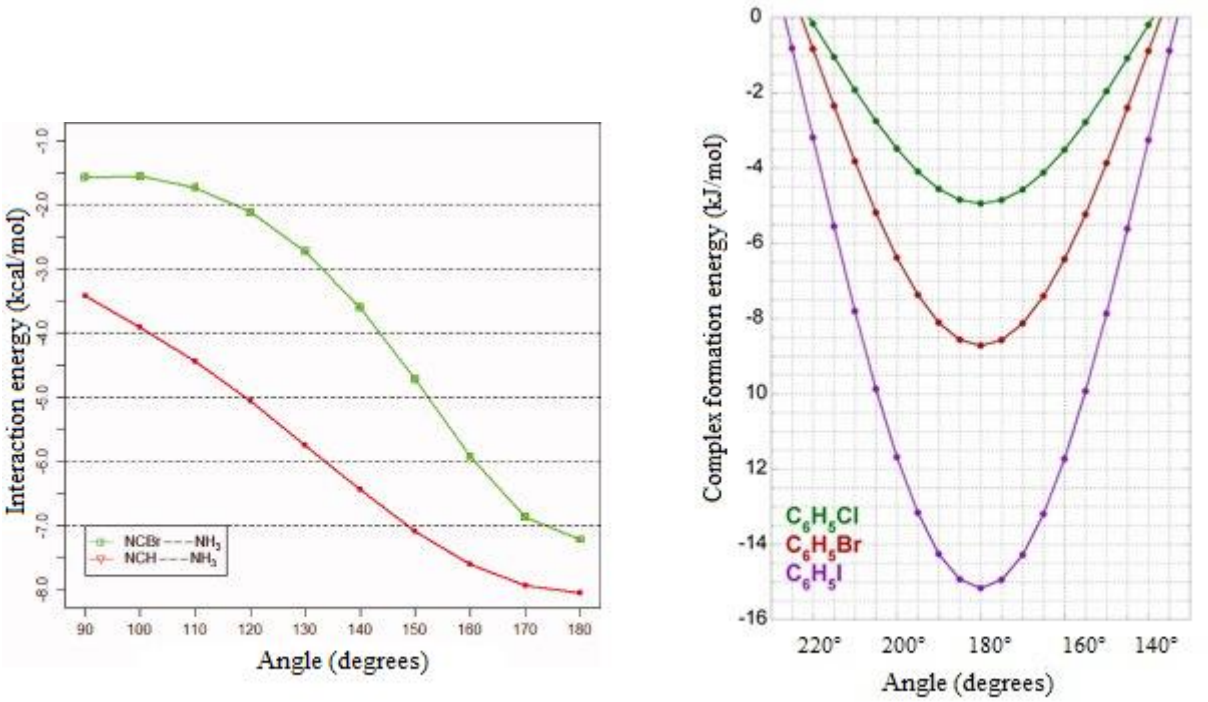


Figure 21. Angular dependence of the strength of halogen and hydrogen bond

The plot at left depicts the dependence of the interaction energy (in kcal/mol) on the angle (the change is only in one direction from the starting orientation). Green line halogen bond, red line hydrogen bond. Adopted from Shields et al. (2010). The plot at right shows a similar angular dependence on the interaction energy - complex formation energy of three halogen bonding compounds complexed with a carbonyl group (in kJ/mol; 1 kcal = 4.2 kJ, the change is in both directions, see Fig. 20). Adopted from Wilcken et al. (2013).

Chalcogen bond can probably tolerate larger deviations from the optimal geometry than halogen bond - halogen bond is more directional (Wang et al., 2015; Pavan et al., 2015). The strong directionality has been often viewed more as an advantage for the drug design, though.

2.3.3.6.5 Sensitivity to solvent

Considering the potential roles of σ-hole bonding (as well as roles of other interactions) in biological systems, all the interactions occur in rather complex environments, the interacting partners are surrounded by water molecules, proteins, etc. Thus approaches based on simple two

interaction partner models in the gas phase are limited. The desolvation penalty of an interaction between two polar residues accounts for the free energy loss given by the removal of the surrounding water molecules from both compounds, so that they can interact with one another (Bissantz et al., 2010). Therefore if a lot of favorable interactions with water molecules has to be lost in order to attain the protein-ligand interaction, such loss disfavors the binding. Desolvation penalties play major roles in interactions between polar or charged compounds. Conversely, interactions between nonpolar compounds, for instance π-π stacking interactions, benefits from desolvation (http://www.cambridgemedchemconsulting.com/resources/solvation.html).

It was proposed that the solvation penalty for halogen bond should be minimal - much smaller than for hydrogen bond, because halogen bond depends on dispersion more than hydrogen bond (Riley and Hobza, 2013). Consistently, it was shown that the strength of selected hydrogen bonds vary with solvent, however sample halogen bonds tend to maintain the order of magnitude of the association strength (Table 3, Fig. 23) (Robertson et al., 2014; Sarwar et al., 2010).

Table 3. Strengths of halogen and hydrogen bonds in various solvents

Depicted values are association constants per mole. Compounds 1-4 are shown in Fig. 22. Chosen solvents from Robertson et al. (2014).

Solvent	Hydrogen bond		Halogen bond	
Interacting compounds (fig.22)	1…3 (M^{-1})	1…4 (M^{-1})	2…3 (M^{-1})	2…4 (M^{-1})
Octane	2400	370	12	8800
Tetrachloromethane	410	24	6	7300
Toluene	230	4	3	11000
Dichloromethane	90	9	2	58000
Chloroform	53	2	1	20000
Dioctyl ether	<1	<1	<1	1600
Acetone	2	<1	<1	1900
Ethanol	<1	<1	<1	3200

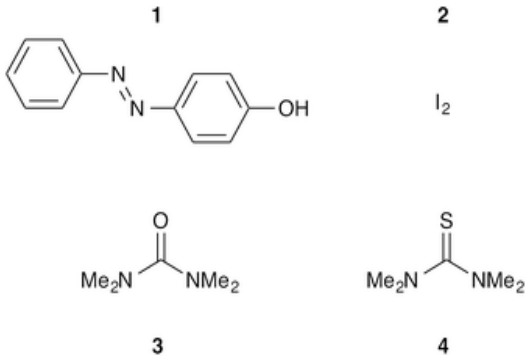


Figure 22. Compounds studied by Robertson et al. (2014)

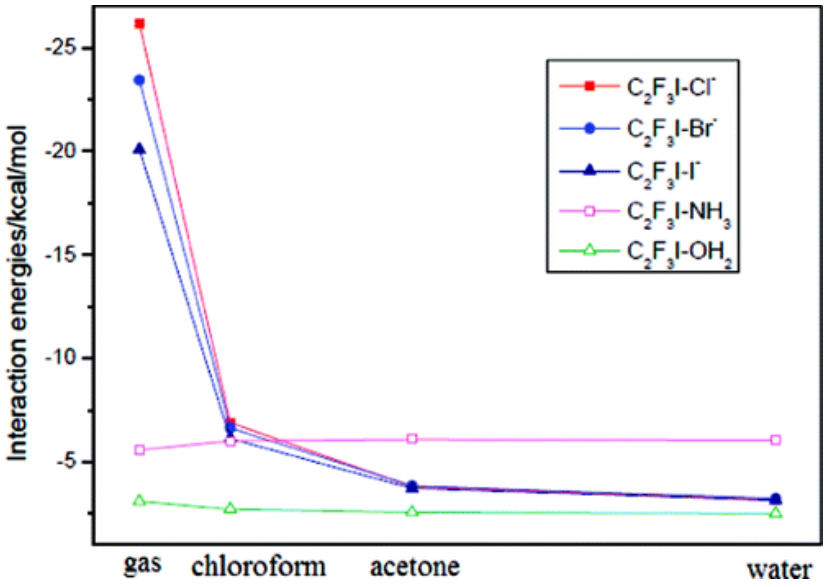


Figure 23. Sensitivity of halogen bond to the solvation

Sensitivity of interaction energies of iodoperfluorethane complexed with various electron donors to the solvation. Red, blue and purple lines are for charged halogen bond complexes, pink and green for neutral (all the halogen bonds discussed thus far). Adopted from Lu et al. (2011).

(Iodoperfluorethane formula should be correctly C_2F_5I, correspondingly with model figures in this publication)

The authors argue that it is the charge transfer that is responsible for such behavior of halogen bond (Robertson et al., 2014). Neutral halogen bonds (Fig. 23, pink and light green) might even be stabilized by solvation (Carlsson et al., 2015). This behavior is obviously advantageous for the rational inhibitor design.

Halogen bond has been found in hydrophobic environment, in the core of a protein, as well as in polar environment, on the protein surface (Hardegger et al., 2011), and in both cases it was important for the interaction.

2.3.3.6.7 Deciphering the physical forces participating in σ-hole interactions

It has been established, that electrostatic force, dispersion force and charge transfer are players in σ-hole interaction. There apparently is a debate about their relative contributions though.

As the theoretical concept behind σ-holes implies (see section 2.3.3.2) electrostatic interaction is an important factor determining the σ-hole bonding (Clark et al., 2007). Indeed the electrostatic interaction is indisputably a very important contribution to the σ-hole bonding (Riley and Hobza, 2008; 2013; Mitoraj and Michalak, 2013; Grabowski, 2013a; Syzgantseva et al., 2013). This force might also be responsible for the directionality of these interactions (Murray et al., 2013; Riley and Hobza, 2008; Legon, 2010). Electrostatics is however not the only force participating in the halogen bond, as can be implied from its behavior in the solvent (Sarwar et al., 2010).

According to some studies, electrostatics with dispersion are predominant in determining the σ-hole bonding (Riley and Hobza 2008; Zierkiewicz et al., 2015) and they are of comparable magnitudes (Riley et al., 2013) (Table 5).

Table 4. Decomposition of the interaction energy of halogen bond

A comparison of the binding energy of four halogen bond complexes and two hydrogen bond complexes interacting with formaldehyde (model systems). Induction term includes charge transfer phenomenon, exchange repulsion is a repulsive component of binding. Interaction energies (Total energy) are in kcal/mol. From Riley and Hobza (2008).

Energy terms	H_3C-Cl	H_3C-Br	FH_2-Cl	F_3C-I	H_3C-H	HCC-H
Electrostatic	**-0.96**	**-1.47**	**-3.11**	**-4.04**	**-0.55**	**-3.78**
Induction	-0.23	-0.37	-0.82	-0.95	-0.16	-0.51
Dispersion	-1.81	-1.98	-2.26	-2.14	-0.72	-1.23
Exchange rep.	2.02	2.12	2.90	2.48	0.72	2.27
Total energy	-0.98	-1.70	-3.29	-4.65	-0.70	-3.25

As it seems, surprisingly, *both* of these contributions get larger when X (halogen bond exhibiting element) goes from lighter to heavier (in the column of the periodic table), however for heavier atoms the electrostatic contribution rises more rapidly than the contribution of dispersion which makes it the major contributor in the interaction (Riley and Hobza, 2008; Syzgantseva et al., 2013). The induction term that include the charge transfer from the electron donor to the antibonding orbital of the halogen was in some theoretical studies (Riley and Hobza, 2008; Wang and Hobza, 2008) the smallest bonding contributor for most of the complexes studied and for all the studied

systems overall (Table 5). Concerning the charge transfer contribution, however, other computational approaches yielded different results, as for the same compounds (as in Table 4) the orbital interaction (charge transfer) with polarization (defined as ”empty/occupied orbital mixing on one fragment due to the presence of another fragment”) were the most important according to Palusiak et al. (2010).

The importance of the partial covalent character was also proposed by other groups based on yet different computational methods - and often performed on different systems (Mitoraj and Michalak, 2013; Sysgantzeva et al, 2013). And it is the charge transfer, the covalent characteristics, that increases with growing σ-hole facilitated by electron withdrawing substituents - according to Ebrahimi et al. (2016). Some groups argued, that the amount of absolute charge transfer to the antibonding σ orbital indeed correlates with the interaction energy supporting the importance of covalent contribution in halogen (σ-hole) bonding (Grabowski, 2013a; Wang et al., 2015). Grabowski also suggested, that the increment in the *s* character of a Y atom in an Y-X (X is a halogen) indirectly proves substantial charge transfer (alongside with a computational analysis) (2013). The existence of charge transfer to mainly (Wang et al., 2004) antibonding σ orbital itself is beyond dispute, only its relative contribution is in question.

The dispersion seems to be a nonsignificant term in the systems studied by Mitoraj and Michalak (2013), and it varies from nonsignificant to very significant system by system in the work of Sysgantzeva et al (2013).

Few studies that examined driving forces of chalcogen bond concluded that charge transfer first accompanied by electrostatics and then dispersion are important (Adhikari and Scheiner, 2014. Charge transfer is the dominating force in halogen and chalcogen bond alike, according to Wang et al., (2016) and is the one factor responsible for the directionality of the chalcogen bond. As in the previous case of halogen bond, other studies suggest, that electrostatics and dispersion are the most important and it is the electrostatics that dominates the directionality (Tsuzuki and Sato, 2013).

These somewhat conflicting results on the determining factors of the σ-hole interactions may arise from the different description of terms for physical forces/effects by the respective methods (Pendás et al., 2006). It was suggested, for example, that charge transfer can be absorbed by the electrostatic term, possibly explaining great deviations presented by different groups (Wang et al.,

2014). This significant mismatch in various methods is a daunting issue to solve - if it is possible. Some researchers even gave up on separating the interaction energy into various terms arguing that such a decomposition is arbitrary, without a rigorous basis and various terms may even not be independent (Politzer et al., 2010). This resignation upon deeper understanding of physical forces behind halogen bond, even though it is understandable, would be an unsatisfying conclusion provided that these contributions are real and at least partly distinguishable entities, as they are generally accepted to be for the case of hydrogen bond.

Conclusions and ideas

Originally, σ-hole interactions were explained by means of empty lobe of one of halogen (or other σ-binding element) *p* orbital (conventionally denoted as p_z) facilitating local positive charge and eventually electrostatic attraction as described in the section 2.3.3.2. Now in frame of energy decomposition schemes new factors are introduced.

The compatibility with charge transfer:

When two elements A and B are bound together by a single bond between their respective overlapping *p* orbitals, both binding orbitals combine into a bonding σ_p and an antibonding σ^*_p. The nonbonding *p* lobe described above is to be understood as an actual nonbonding lobe of a σ_p molecular orbital where the antibonding orbital σ^*_p is a different entity. In this concept bonding and antibonding orbitals are separate entities, even though the space they occupy somewhat overlaps (Fig. 2). In σ-hole exhibiting compounds, the lack of electronic charge on σ_p attracts electrostatically with electron rich moieties on the interaction partner. Then also the lack of electronic charge around σ_p might "unmask" an empty antibonding σ to which some portion of the electronic charge from an approaching electron donor C might be (and apparently is) transferred.

Inclusion of dispersion:

If we expand the concept of dispersion on charged or partially charged compounds, then inclusion of dispersion would be compatible with the original σ-hole interpretation presented by Clark et al. (2007). Electron clouds on the polar regions of compounds (on a σ-hole and its interacting partner, nucleophile) can also fluctuate and their fluctuation can be aligned to produce additional attractive force. When both interacting compounds are

permitted to be very close due to the electrostatic interaction as is the case for σ-hole interactions, short ranged dispersion becomes increasingly relevant.

Polarization:

The electron density in the molecular orbital corresponding to nonbonding p_z orbital in σ-hole binding compounds can be further polarized towards electronegative substituents, increasing the σ-hole. Also the electronic density of the overall electron cloud around the σ-hole binding compound can be polarized away from σ-hole that again works for its increase.

Aims of the Thesis

The aims of this work are following:

1) To investigate the occurrence of chalcogen bonds in viral protein-ligand complexes in the PDB database.
2) To explore the strength and geometrical preferences of found chalcogen bonds in model complexes.
3) To optimize the chalcogen bonds strength in the model systems.

3. Materials and Methods

3.1 Choice and elaboration of model systems

Proteins of interest have been chosen from PDB (Protein data bank, http://www.rcsb.org/pdb/home/home.do) by a manual search among structures with criteria: resolution below 2.01 Å, distance between a chalcogen atom (sulfur, selenium or tellurium) of a ligand and a nucleophile of a viral protein below or close to the sum of van der Waals radius. Other than divalent chalcogen containing compounds were excluded from the search since there is no evidence in the literature that these compounds could participate in σ-hole bonding. Additionally, online resources for sequence comparison (BLAST similarity search using FASTA sequences)

were used to examine the homology of selenium containing ligand interacting with proteins of nonviral origin and viral proteins.

Protein data bank

The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) founded in 1971 holds information of experimentally determined protein and nucleic acid structures, including complexes with ligands. Today the databank contains over 100 000 macromolecular structures designated by four character code (PDB ID). PDB is also associated with options of online query for structures of desirable parameters and tools for sequence comparison (Bourne et al., 2011).

Elaboration of model structures

Model systems containing presumed interaction were prepared from these proteins using Pymol 1.5.0.4 (https://www.pymol.org/), software for molecular visualization which is also capable of refinement and building of molecules.

3.2 Quantum mechanical calculations

For all the calculations, in-house interface Cuby4 (http://cuby4.molecular.cz/) was used. The Cuby software is capable of employing different softwares for quantum and molecular mechanics calculations and combine their results.

4.2.1 Electrostatic potentials

Electrostatic potentials of separated interaction partners were calculated by the Gaussian 09 software (http://www.gaussian.com/g_prod/g09.htm) using the Hartree-Fock (HF) method with the cc-pVDZ basis set, to be consistent with the literature (Politzer et al., 2013). The visualization was performed in the Molekel 4.3 software (http://molekel.software.informer.com/) on 0.001 a.u. isosurface (electrons per $Bohr^3$) that contains 96% of molecule´s electric charge defined by Bader et al. (1987).

Hartree-Fock calculations

Hartree-Fock or self-consistent field method is a wave function-based approach of solving the Schrödinger equation. It is based on the Born-Oppenheimer´s approximation, which

separates the effect of the nuclear motion and electronic motion. The HF approach also takes all the electron-electron interactions in an average fashion (Jensen, 2007).

4.2.2 Interaction energy computations and gradient optimizations

The interaction energy is calculated as a difference between the energy of a complex (AB) and the energy of both individual compounds (A) and (B). Thus the interaction energy ΔE of a complex AB is defined as $\Delta E_{(AB)} = E_{(AB)} - E_{(A)} - E_{(B)}$. The more negative the interaction energy, the stronger the binding is.

The density functional theory (DFT) with a parametrized dispersion (D3) was used with the TPSS functional and the TZVPP basis set as available in the Turbomole software, ver. 7 (http://www.turbomole.com/). The convergence criteria for free optimizations were ΔE limit 0.0006 kcal/mol and the maximum gradient limit 0.12 kcal/mol·Å. Criteria for other optimizations (optimization of hydrogens or optimization utilized for angular scan) were ΔE limit 0.006 kcal/mol and the maximum gradient limit for these cases was 1.2 kcal/mol·Å. For distance and some of angular scans (5.3.1.6.2, 5.3.1.6.3, 5.3.2.6.2, 5.3.2.6.3) the systems were prepared using the Molden software 5.4 (http://www.cmbi.ru.nl/molden/).

Density functional theory

The density functional method (DFT) is not based on wave-function approaches such as HF methods, but instead on the Hohenberg and Kohn statement. The statement says that the ground state energy of molecular system is completely described by its electron density. Instead of working with separate wave functions of electrons (as HF-based methods do) where the complexity increases with the increasing number of electrons, the DFT approach that utilizes electronic density has the same number of variables for small and larger systems alike and is thus much more efficient than wave function-based methods. The energy in the DFT methods is divided into the electronic kinetic energy, the attraction between the nucleus and electrons and the electron-electron repulsion. One of the DFT´s flaws is its limited capability to describe weak dispersion force driven interactions as it does not describe electron correlation properly (especially at longer distances). Empirical correction terms for dispersion are added for this reason, for example D3. The D3 adds a dispersion correction term that relies on empirical dispersion coefficients for atom pairs (Jensen, 2007; Grimme, 2011).

3.3 Strategy of calculations

To study interactions of interest efficiently, chosen protein structures were simplified to model systems, which contained two interacting partners only – a representative of the ligand and of the protein derived moiety (side chain or carbonyl representative), omitting the rest of the molecule. Doing so made the computation faster and did isolate the interacting partners from possible secondary interactions that may have caused the results to be difficult to interpret. On the other hand reducing the complexity of the model systems to extent greater than necessary may impair the plausibility of the prepared model system. This could be caused by the neglection of possible intramolecular contributions of not included functional groups to the interaction of interest (for example caused by exclusion of the interacting group polarization effect by not included moiety). Thus it is important to choose optimal model systems and also to keep in mind that other effect may influence the interaction of interest in proteins. Model systems were elaborated from the ligand by keeping the smallest model possible while respecting aromatic rings and residues that might be affecting the σ-hole´s magnitude (electron withdrawing or donating groups) that are either close to the chalcogen or on the same aromatic ring as the chalcogen. Hydrogen atoms not described by the crystal structure were also added in the model system elaboration process.

For the σ-hole tuning the model systems were modified so that the σ-hole would be more significant. This was done correspondingly to the findings of other researchers mentioned in the section 2. Literature Review. Carbon bound hydrogen atoms were replaced with fluorine atoms and ring sulfur heteroatoms were replaced with selenium heteroatoms in these systems. This system subset will be referred to as “substituted systems” further on.

For crystal structure geometries all the added atoms (hydrogen atoms and for substituted systems hydrogen, fluorine and selenium atoms) were optimized first correcting possible imperfections of an automatic atom adding. Other atoms were frozen to keep the systems most analogous to the experimental input. After that, for each system the full gradient optimization was performed, finding the local energy minima. From that point, the restrained optimization was done, forcing the interacting partners to align in the orientation where the σ-hole would theoretically be the most exposed to the interacting nucleophile and the part of chalcogen bond on the total interaction energy would be the greatest. This step would help the systems to reach the energy minimum connected

with the chalcogen bond formation in case of the free optimization procedure before would not locate it. Subsequently, systems were freely optimized again without any restraints, finding local minima or confirming chalcogen bond minima found in the previous step. The systems that would deviate too much from the theoretical chalcogen bond optimum would be assumed that σ-hole plays no important role in their binding. The systems that did not maintain the presumed σ-hole binding motif at least partially were excluded from the further examination. For the rest, angular and distance scans were performed. In frame of angular scans the interacting partners were shifted in the principal x, y and z planes preserving certain interactions and eliminating others in order to grasp an effect of the σ-hole bonding. Angular scans were done in steps of 10° from the starting geometry - the geometry of the systems after second optimization. Distance scans did measure the change of the interaction energy with the increasing separation of the interacting partners and were done by steps by 10% of the sum of respective atoms van der Waals radii.

4. Results

4.1 Protein data bank search

The PDB database displays as many as 11,795 protein-ligand structures below 2.01 Å resolution, 9,592 proteins of a viral origin and 4,199 proteins containing ligand with a sulfur atom, 60 containing selenium ligand and only 5 containing tellurium ligand. With all of these criteria combined, the PDB database returned 217 entries for protein-ligand structures. No entries have been found with a selenium or tellurium atom containing compound, nor have been found nonviral selenium (or tellurium) ligand containing compound homologous to any viral protein (data not shown). From these, only four ligands containing (divalent) sulfur were found to presumably display σ-hole bonding (Table 6). These chosen crystal structures are all of proteins of medically relevant viruses. For comparison, there are another 185 results for ligands of viral proteins containing sulfur that have resolution between 2.01 and 3.01 Å.

4.2 Studied crystal structures

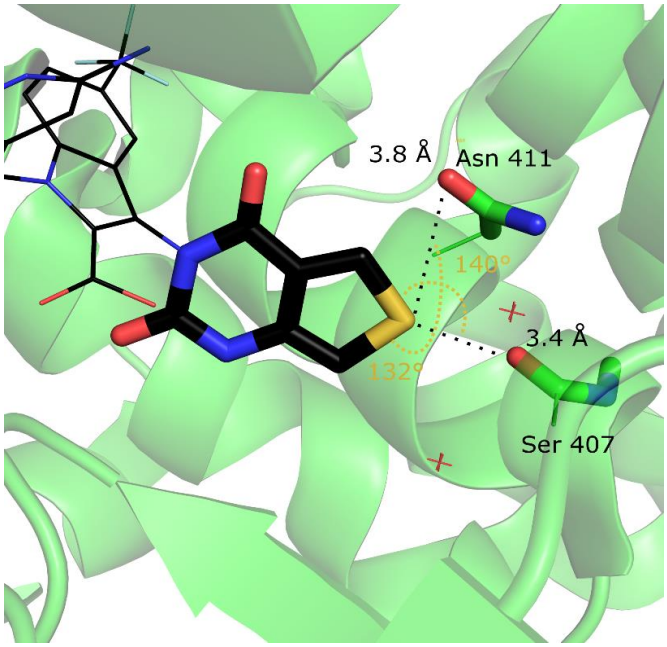
Table. 5. Selected crystal structures

Crystal structures of viral proteins with a sulfur containing ligand satisfying the resolution criteria found in PDB that might exhibit a chalcogen bond. In all the systems the sulfur to oxygen atom interaction motif was observed. 3SKE systems might also exhibit a C-N...O motif, after a slight accommodation in the binding site.

PDB code	Resolution (Å)	Organism	Protein	Interaction motif
3SKE	1.97	Hepatitis C virus	NS5B Polymerase	C-S…O (C-S…N)
4CTK	1.53	Dengue virus	Methyl transferase	C-S…O
4IBK	1.85	Ebola virus	VP-35	C-S…O
4U7Q	1.70	HIV	Protease	C-S…O

5.2.1 3SKE

PDB entry 3SKE is of the hepatitis C virus NS5B RNA-dependent RNA polymerase. NS5B is a homodimer with two active sites both being occupied by a ligand (Fig. 24). The NS5B monomer has a finger, thumb and a palm domain. The active site is situated in a cleft between those three domains. The 3SKE ligand is the second best in terms of of IC_{50} and EC_{50} from a series of ligands designed to rationally optimize initial "lead" inhibitor (Anilkumar et al., 2011). Residues involved in catalysis lie within the palm domain. Aspartate 220, 318 and 319 bind divalent metal ions stabilizing nucleotides in the active site and additional residues R158, S367, R386, T390 and R394 contact the incoming nucleotide (Ranjith-Kumar and Kao, 2006). The sulfur atom of thienopyrimidinedione ring in chain A is 3.4 Å (3.5 for the chain B) away from a carbonyl oxygen – roughly at the sum of van der Waals radii that is 3.32 Å (Fig. 24). The C-S…O angle of 140° (141° for chain B) was less suitable for a σ-hole interaction, though. Asparagine 411 has even less suitable orientation for a possible chalcogen bond (S…O distance 3.8 Å and C-S…O angle 132° or S…N 3.8 Å and C-S…N 155° in the case that the asparagine nitrogen atom would be an



interacting nucleophile). However the ligand could theoretically rearrange to a σ-hole permitting orientation by a simple rotation along the C-N bond linking the N-thienopyrimidinedione ring to the rest of the inhibitor. This would probably happen if the σ-hole had more positive ESP. The ligand also contains one carboxyl group that is exposed to the solvent, the ligand as a whole is probably an anion unless it is protonated.

Figure 24. 3SKE ligand bound to NS5B polymerase of HCV

The 3SKE ligand in an active site of a protein, chain A is shown. Only the considered residues asparagine 411 and serine 407 are shown. The ligand´s sulfur atom in the aromatic ring is close (in terms of van der Waals radii) to a carbonyl oxygen of the serine 407 amino acid. The sulfur atom is further away from an oxygen atom of asparagine 411 side chain. Carbon atoms of the ligand are in black, carbon atoms and ribbons of protein are green. Model systems further studied are shown by thick lines. Color and thickness designation applies to the other systems (below).

From 3SKE, model systems for both possible interactions were chosen. The thienopyrimidinedione ring was chosen to represent the ligand (Fig. 24, carbon atoms in black, thick lines) and was complexed with either N-methyl formamide representing a carbonyl backbone - 3SKE (1) or carboxamide representing the asparagine side chain - 3SKE (2) (carbon atoms in green, thick lines). (N-methyl formamide and carboxamide are representatives of asparagine side chain or carbonyl backbone respectively in all the systems). Model systems of 3SKE (3) and (4) correspond to (1) and (2) respectively, only they have thiophene ring instead of the whole thienopyrimidinedione as a ligand model. Substituted systems 3SKE (A) and 3SKE (B) correspond to 3SKE (1) and (2) respectively, and do account for model systems tuned with carbon bound hydrogen to fluorine and sulfur to selenium atoms substitutions.

Further model systems designated with letters in brackets are referred to as substituted systems and are optimized to display greater σ-holes similarly to 3SKE (A) and (B). Model systems designated with numbers in brackets are unsubstituted and referred to as crystal structure model systems or unsubstituted models. It is however apparent that in the original 3SKE crystal structure the interacting partners are nowhere near the optimal geometry that would be anticipated for an interaction in which the σ-hole would be a dominant contributor. It would therefore be necessary

for the optimized compounds to accommodate in the active site such that the optimized σ-holes would be exploited to the maximum.

5.2.2 4CTK

4CTK is the homodimeric dengue virus methyltransferase that serves for the formation of a 7-methylguanosine cap structure. The dengue virus methyltransferase consists of three distinct binding sites. One that binds to adenosylmethionine or adenosylhomocysteine, which serves as a methyl group donor in the methyltransferase reaction, GTP binding site that recognizes the cap structure and a positively charged RNA binding site. 4CTK protein´s adenosylmethionine binding pocket is bound to two molecules of S-adenosylmethionine. In chain A, thieno[2,3-b]pyrazin-7-amine binds to the GTP binding pocket of the enzyme in two different orientations (Fig. 25). The ligand in crystal structure 4CTK is one of the candidate structures chosen by *in silico* screening targeting dengue methyltransferase. The 4CTK ligand is capable of a slight inhibition of 7-guanosine methyltransferase activity (9%) and even more negligible inhibition of 2´ ribose methyltrasferase activity (Coutard et al., 2014). The binding of this ligand is probably stabilized by an obvious π-π stacking interaction to the phenylalanine 25 aromatic system. Aside from this stacking interaction there is a ligand sulfur atom to leucine 17 carbonyl oxygen atom close contact with sulfur to oxygen distance of 3.1 Å. With such an interatomic distance below the sum of van der Waals radii and with an angle C-S…O of 165° there is a possibility for a chalcogen bond to participate in the binding. The ligand is forming few contacts with adjacent protein residues – a rather simple system for further optimization. However the aromatic system stacking with an adjacent phenyl group would keep the ligand ring in a fixed plane, thus greater deviation from this plane to accommodate to the σ-hole would be probably accompanied with an energy penalty. There is another ligand molecule below the described ligand completing a triple sandwich stabilized by a π-π stacking interaction.

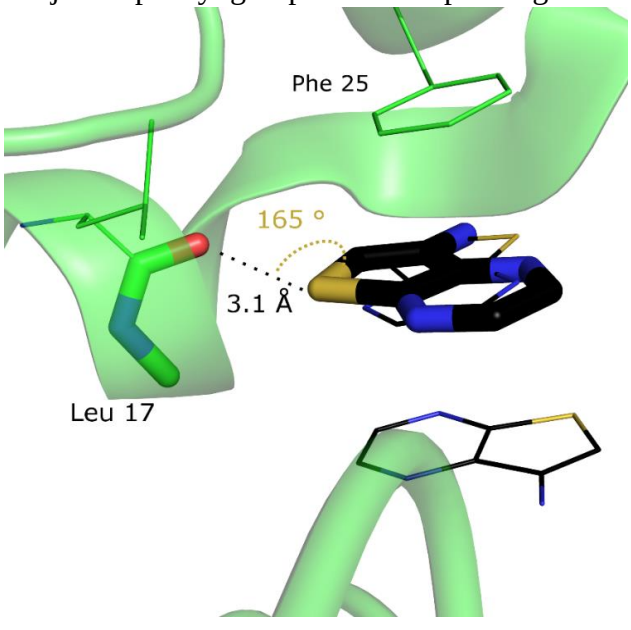


Figure 25. 4CTK ligand bound to dengue virus methyl transferase

The 4CTK ligand bound to the GTP binding pocket of a chain A. A short contact of an aromatic ring sulfur with a leucine 17 carbonyl oxygen along with a suitable geometry for a chalcogen bond is apparent. A stacking interaction

with a phenylalanine 25 ring above probably facilitates the binding. Another molecule of the ligand is situated below creating triple sandwich of aromatic rings stabilized with a π-π stacking interaction.

For the ligand models – 4CTK (1) as well as for the substituted system 4CTK (A) the ligand was kept intact as the aromatic rings (or double bonds) can allow stronger inductive and mesomeric effects as the delocalized electrons can more easily move in a molecule.

5.2.3 4IBK

4IBK is a crystal structure of small homodimeric Zaire ebolavirus VP35 protein complexed with a ligand. VP35 is a multifunctional protein that serves as a viral polymerase cofactor and it also inhibits host interferon production. VP35 consists of N-terminal coiled-coil domain, C terminal domain that contains a site important for ds-RNA binding and interferon inhibition and a site important for its polymerase cofactor function. 4IBK ligand was one of compounds chosen from a large set of possible ligands by *in silico* screening and exhibited reasonable value K_D of about 50 μM (Brown et al., 2014). The sulfur atom of the chlorothiophene ring is 3.5 Å away from the oxygen atom of glutamine 241 side chain (in both chains) and the angle of C-S…O is 151° (148° for chain A) (Fig. 26). The ligand contains one carboxyl group, thus is negatively charged. The ligand´s carboxyl group is forming a salt bridge with a side chain of lysine 251.

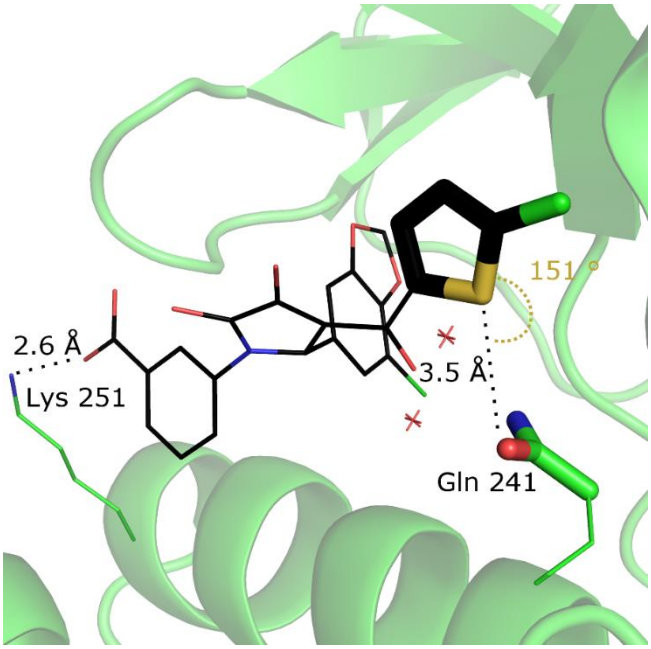


Figure 26. 4IBK ligand bound to ebola virus VP35

The sulfur atom of chlorothiophene ring is close to the glutamine 241 oxygen atom, with an orientation not completely excluding a chalcogen bond. The 4IBK ligand is an anion and its carboxyl group interacts with a lysine 251.

The model for the ligand was chosen to be chlorothiophene – 4IBK (1). To investigate the role of a chlorine atom, unsubstituted thiophene-carboxamidee complex model system was also prepared - 4IBK (2). Substituted system with sulfur replaced with selenium atom and all the ligand hydrogen atoms (and single chlorine) with fluorines was prepared, too - 4IBK (A).

5.2.4 4U7Q

4U7Q is a crystal structure of homodimeric aspartic protease of HIV 1. The HIV protease cleaves the viral GAG-POL polyprotein mediating the maturation of newly assembled particles. In a cavity between the two 99 amino acid long chains lies a catalytic site. Aspartate 25 of both chain facilitate the hydrolysis of a peptide bond. The central cavity is protected by two flexible segments (flaps) that open to permit the polyprotein substrate enter and the product leave the active site. The 4U7Q ligand is an inhibitor that binds to the active site in two alterative orientations with relative occupancies of 0.5 (Fig. 27). The ligand exhibits about 170 pM inhibition constant K_i and is photodegradable. The photodegradation reduces the inhibitory activity by about 4 orders of magnitude (Schimer et al., 2015). The thiazole rings in chains A and B are situated 2.5 and 3.2 Å respectively from the glycine 48 main chain carbonyl oxygen atom. While the glycine 48 A chain interacting thiazole ring being only 2.5 Å away from the oxygen atom, that is only 75% of the sum of van der Waals radii, it is inappropriately oriented to display any σ-hole bonding. From this respect the chain B is much more promising. On the other hand if the ligand is able to adopt such orientation suitable for a chalcogen bond as in chain B, it might accommodate to the most σ-hole favorable geometry after further ligand tuning easily.

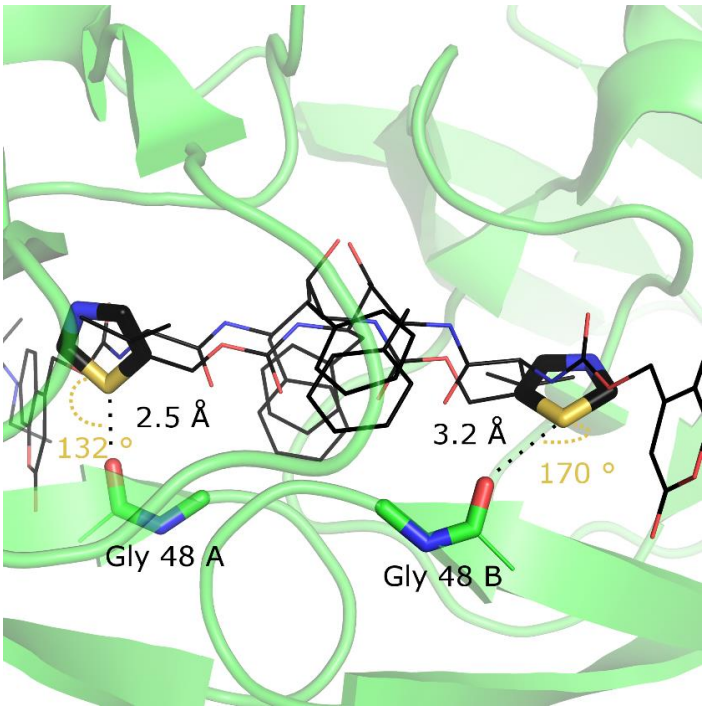


Figure 27. 4U7Q bound to the HIV protease

The ligand is bound in the active site in two alternative orientations due to the symmetry of the HIV protease. In both cases ligand´s thiazole ring sulfur atom is very close to a glycine 48 carbonyl oxygen atom of either chain A or B. While in A chain sulfur and oxygen atoms are much closer, its orientation is not particularly favorable for a σ-hole interaction. In chain B the thiazole ring is in promising orientation to the carbonyl oxygen considering the chalcogen bond possibility.

The thiazole ring represented the ligand in model system. Thiazole rings in both chains were examined for comparison of alternative starting geometries - 4U7Q (1) and (2). Of substituted systems, 4U7Q (A) and (B) correspond to (1) and (2), respectively.

5.3 Model systems

5.3.1 Model systems derived from crystal structure

5.3.1.1 Electrostatic potentials (ESPs)

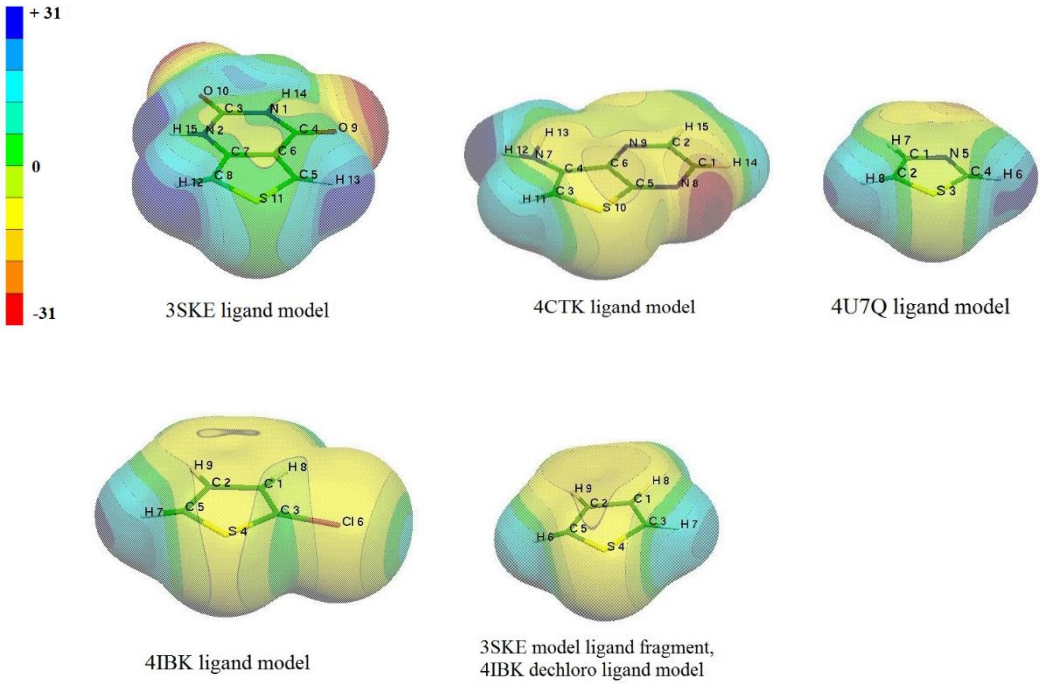


Figure 28. ESPs of ligand models

For all the models the same color scale of ESP from +31 to -31 kcal/mol is used (for amino acid models in Fig. 29, too). Areas inclining to blue shades are positive, areas inclining to red shades are negative. 3SKE model ligand fragment corresponds to 3SKE (3), (4) and also 4IBK (2) model system ligands.

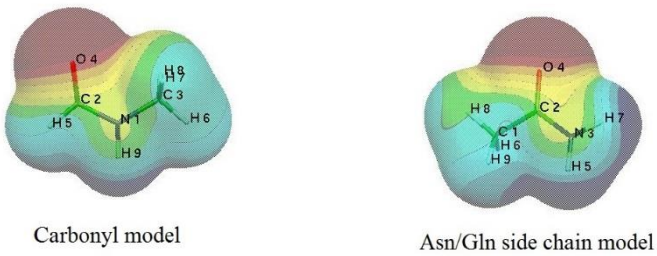


Figure 29. ESPs of amino acid ligands

ESP´s of the carbonyl model (left) and the asparagine/glutamine side chain model (right) are similar with the respect to partially negative moieties (oxygen atoms) interacting with presumed σ-holes of ligand models.

5.3.1.1.1 Electrostatic potentials of ligand models

σ-holes are best apparent in case of 3SKE ligand models (1) and (2) (see Table. 7.). In 3SKE (1) and (3) the nucleophile in the original crystal is closer to a hydrogen atom H 12 (Fig. 28) site of the ligand. Thus this site (and adjacent σ-hole) is considered as the interacting site in further steps for these model systems. Electrostatic potential on this side is more positive for 3SKE (1), but also nearby hydrogen atom (H 12) is more positive than the σ-hole itself. Interestingly the noninteracting σ-hole of a 3SKE (1) ligand model is more positive than a nearby hydrogen (H 13). This site, where the σ-hole is more positive than nearby hydrogen, is considered as an interacting site for 3SKE (2), as it is closer to its presumed interaction partner. This site is considered as an interacting one for 3SKE (4) as well, but this model ligand does not have an apparent σ-hole.

4CTK (1) model also has a distinct σ-hole that is to some degree blending with the positive areas of a nearby hydrogen atom as for 3SKE (1). In these cases the area around the hydrogen is also slightly more positive than that of the σ-hole. On the noninteracting face of the sulfur atom in 4CTK (1), the entire σ-hole region is negative overall. However there is a distinct σ-hole region relatively less negative than its environment.

Five-membered rings of 4IBK (1), 4U7Q (1) and (2) have also apparent σ-hole areas as protrusions of positive ESP from the nearby hydrogen atoms. 4IBK (1) has a noninteracting σ-hole with a smaller magnitude that is deformed by a close chlorine atom, which has a σ-hole of its own (with the magnitude of about 10 kcal/mol), facing to the solvent. The smallest σ-hole areas are observed in the thiophene ring of 3SKE (3), (4) and 4IBK (2).

In cases of all the studied systems, σ-holes were centered slightly away from their theoretical position. They are (still) in the plane of the aromatic ring, however moved by about 10-15° off from a C_5-S_{11} (Fig. 28, 3SKE as an example) axis towards the C_8 bonding carbon. The C_5-S_{11} axis corresponds to the vector of the nonbonding p_z orbital on which the ESP should be the most positive, as it lacks electrons.

5.3.1.1.2 Electrostatic potentials of amino acid models

ESPs of amino acid models, model of carbonyl backbone and glutamine/asparagine side chain, were comparable one to another, at least in the area around the nucleophilic center (oxygen atom). The magnitude of ESP around the nitrogen atom suggests that it would be a much weaker chalcogen bond acceptor than the oxygen atom. Thereby further on, only oxygen atom will be examined as a possible nucleophile.

Table 6. ESP magnitudes of ligand models

Values of electrostatic potential are in kcal/mol. Magnitudes are measured at centers of circular areas of positive ESP on the sulfur atom (not exactly in the axis of the C-S bond). Interacting site is considered the one of the two σ-holes that is in that concrete model system closest to the interacting partner. For symmetric molecules they are considered being the same and minor deviations in the measurement for both sites (not listed) are attributed to the experimental error. Only the noninteracting σ-hole of 4CTK (1) is slightly negative.

Models	Magnitude (kcal/mol) Interacting site	Magnitude (kcal/mol) Noninteracting site
3SKE(1)	32.00	30.93
3SKE(2)	30.93	32.00
4CTK(1)	21.02	-2.35
4IBK(1)	24.13	4.39
4U7Q(1,2)	22.75	27.30
4IBK(2), 3SKE(3,4)	17.88	17.88

5.3.1.2 Crystal structure geometries of the model systems

After the model systems were elaborated from the crystal structure, hydrogen and halogen atoms were optimized independently prior to the further steps. This corresponded only to a minor movement of these singly bonded compounds fixing possible imperfections of the automatic filling of hydrogen atoms. Resulting model systems were most authentic to the original crystal (including

possible refinement errors). Table 8 shows interaction energies after the optimization of hydrogen and halogen atoms, C-S…O angles and S…O distances of respective compounds.

The best starting orientation for a possible chalcogen bond have 4CTK (1) and 4U7Q (2). From these 4CTK (1) has very strong interaction, about -4 kcal/mol. Its sulfur to oxygen atom distance is well below the sum of van der Waals radii (which is 3.32 Å). In case of 4U7Q (1), where the S...O distance in the protein crystal was too short, the interaction energy is very high and positive (about +9 kcal/mol), the repulsion force is dominant for this system in the current orientation. In case of 4U7Q (2), which is an alternative binding mode for 4U7Q ligand, the geometry around the sulfur atom is reasonable in terms of its plausibility for a chalcogen bond. Its interaction energy is rather medium, though. Both 4IBK model systems (1) and (2) exhibited orientations in which partial chalcogen bond contribution to the binding is still possible, but interaction energies for these systems were weak (about -1 kcal/mol). Interestingly, while interaction energies of 3SKE (1) and corresponding 3SKE (3) differ modestly, interaction energies of 3SKE (2) and (4) did not, even though the respective model system pairs, (1) with (3) and (2) with (4) had the same starting geometry.

Table 7. Parameters of model system after the optimization of hydrogens and after the free optimization

The table depicts characteristic distances and angles in the model systems relevant for the respective chalcogen bonds before and after the optimization. Angles are in degrees, distance in ångströms. For geometries of model systems after optimization see Fig. 30.

Crystal crystal structure geometry				Free optimization		
System	S...O distance (Å)	C-S…O angle (°)	ΔE (kcal/mol)	S...O distance (Å)	C-S…O angle (°)	ΔE (kcal/mol)
3SKE (1)	3.45	140°	-1.52	3.78	151°	-4.85
3SKE (2)	3.76	131°	-2.57	3.30	162°	-4.87
3SKE (3)	3.45	140°	-0.29	4.06	143°	-3.52
3SKE (4)	3.76	131°	-2.60	3.97	148°	-3.98
4CTK (1)	3.15	165°	-4.14	4.01	146°	-4.74
4IBK (1)	3.54	151°	-0.83	4.14	135°	-5.12
4IBK (2)	3.54	151°	-0.70	4.19	129°	-1.84
4U7Q (1)	2.49	132°	+8.89	3.62	135°	-1.07
4U7Q (2)	3.23	170°	-1.35	4.42	130°	-3.71

5.3.1.3 Free optimization

In free optimization step model systems were optimized entirely (all of the atoms), finding the local minima of interaction energy (Fig. 30).

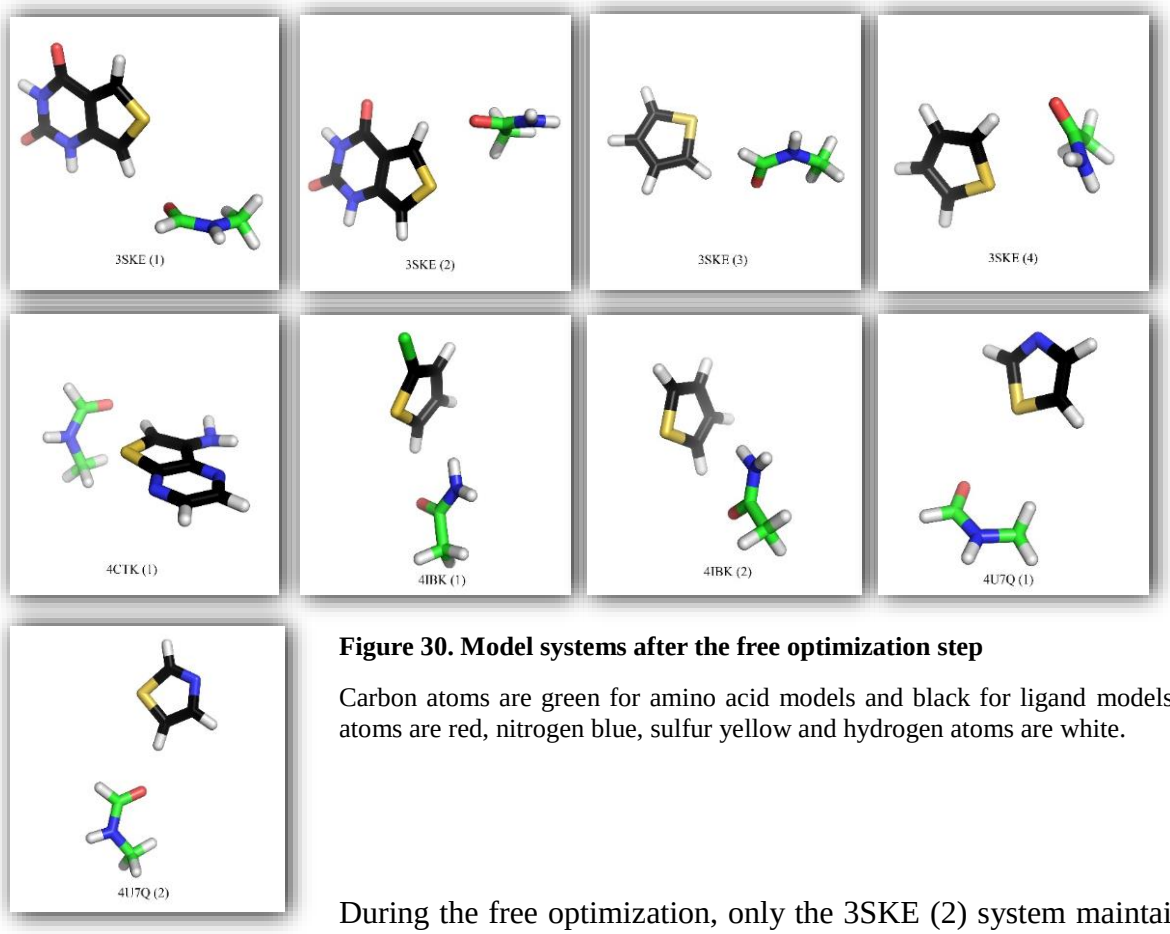


Figure 30. Model systems after the free optimization step

Carbon atoms are green for amino acid models and black for ligand models, oxygen atoms are red, nitrogen blue, sulfur yellow and hydrogen atoms are white.

During the free optimization, only the 3SKE (2) system maintained the S...O distance below the sum of van der Waals radii with a reasonable chalcogen bond permitting angle C-S...O. No other system ended up with the sulfur to oxygen atom distance below the sum of van der Waals radii (Table 8). The interaction for 3SKE (2) after this step was very strong (the interaction energy of about -5 kcal/mol) and the C-S…O angle was the most chalcogen bond favoring (close to linear) among systems. 3SKE (1), and 4CTK (1) maintained a geometry, where the chalcogen bond could play a role. Other systems ended up with orientations less reasonable for the chalcogen bond and their interaction strengths were weaker, except for 4IBK (1), which had the most negative interaction energy among the systems after this optimization step (over -5 kcal/mol). In 4U7Q (1) the interaction energy dropped from almost +9 kcal/mol (repulsion) to values more fitting an actual interaction - weak one, though (about -1 kcal/mol). 4IBK and 4U7Q

systems had the C-S…O angle below 140°, this angle is greater (see Table 8) in 3SKE (3) - the same system as 4IBK (2), only with a different starting orientation.

5.3.1.4 Presumed optimal chalcogen bond geometry optimization

The systems in this step were forced to adopt the linear C-X…O angle, where the σ-holes should theoretically have the greatest contribution to the binding (Table 8). This restrained optimization should help the systems to locate the chalcogen bond featuring energy minimum for cases where the free optimization procedure (section 5.3.1.3) could not reach it because of energy barriers.

After the C-S...O angle being forced to 180°, 3SKE (1), (2), 4CTK (1) and 4U7Q (2) adopted sulfur to oxygen atom distance below the sum of van der Waals radii, while only in the system 3SKE (2) was the S...O distance so short before. The strongest interaction was observed in the model systems 3SKE (1) and (2) (about -4 kcal/mol). The interaction energy was for each system except 4U7Q (1) less negative than in the previous step, though. Thereby, those model systems have left the local energy minimum.

Table 8. Interaction energies and S…O distances after the optimization into the presumed chalcogen bond optimal geometry

Interaction energies (ΔE) of the systems after the C-X...O angle was forced to 180° are in kcal/mol. Distances are in ångströms.

C-S...O angle to 180°		
System	ΔE (kcal/mol)	S...O dist. (Å)
3SKE (1)	-3.95	3.11
3SKE (2)	-3.59	3.10
3SKE (3)	-1.68	3.61
3SKE (4)	-2.88	3.52
4CTK (1)	-3.31	3.13
4IBK (1)	-2.48	3.67
4IBK (2)	-1.84	3.95
4U7Q (1)	-1.65	4.00
4U7Q (2)	-2.25	3.23

5.3.1.5 Free optimization from the chalcogen bond optimal geometry

There was a question whether the geometry forced in the previous step lies in the energy minimum or how far from it is from the minimum. To answer this question, another free optimization was performed for model systems where the starting point was the geometry with the linear C-S…O angle.

Interacting partners deviated from the theoretical C-S…O angle of 180° in all the systems (they were not in energy minimum). 4IBK (2), 4U7Q (1) and (2) did deviate too much from the expected orientation, and thereby will not be examined further as there probably is a small chalcogen bond contribution in this system´s interaction energy. The second free optimization step was connected with the interaction strengthening in all the systems. Interaction energies and geometries after this step were very similar to those after the first optimization for most of the systems. Only 4U7Q (1) found a new local energy minimum (-3.52 kcal/mol). In no system did the nucleophile get too far away from the ring plane, however out of the studied systems only 3SKE (2) ended up with a reasonable chalcogen bond favoring geometry (with the angle C-S...O of 165°) – about where it was after the first optimization step. From 4IBK systems, system (2) without a chlorine moiety optimized roughly to an orientation with an oxygen atom facing the C-H bond (away from the sulfur atom), whereas 4IBK (1), that is a ligand model substituted with a chlorine atom, ended up with the geometry possibly permitting some chalcogen bond contribution after this step. These two systems had roughly the same interaction energy of about -5 kcal/mol, though – the most negative among the systems.

Table 9. Geometrical parameters and interaction energies (ΔE) of model systems after the second free optimization

Interaction energies (ΔE) are in kcal/mol, distances are in ångströms and angle is in degrees.

System	S…O Distance (Å)	C-S…O angle (°)	ΔE (kcal/mol)
3SKE (1)	3.83	149°	-4.85
3SKE (2)	3.23	165°	-4.80
3SKE (3)	3.90	147°	-3.63
3SKE (4)	4.02	149°	-3.99
4CTK (1)	4.13	143°	-4.96
4IBK (1)	3.81	146°	-5.04

4IBK (2)	4.44	125°	-4.99
4U7Q (1)	4.25	136°	-3.52
4U7Q (2)	4.16	130°	-3.75

5.3.1.6 Scans

Three kinds of angular scans and one distance scan were performed to ascertain the change of interaction energy with a controlled change of the geometry. The starting orientation was the one attained by the free optimization from the chalcogen bond optimal geometry (section 5.3.1.5).

5.3.1.6.1 Angular scan 1

(Out of the plane of the aromatic ring)

This scan based on the chalcogen bond directionality was meant to isolate the respective chalcogen bonds from other interactions by turning the ligand molecule around the nucleophile (turning the aromatic ring over). Distances of chalcogen to oxygen (chalcogen bond) and oxygen to hydrogen (secondary interaction – hydrogen bond) were maintained constant (while the ring turning, sulfur atom and interacting hydrogen atom were kept frozen) (Fig. 31). The resulting interaction energy changes should be only due to the chalcogen bond (and possibly to lesser extend also the hydrogen bond) directional preferences.

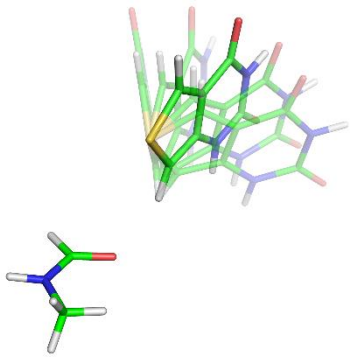


Figure 31. Scan 1, from the plane of an aromatic ring

The angular scan 1 was performed with the ligand model turning around the oxygen of an amino acid model with sulfur to oxygen and also ligand hydrogen to oxygen atom distances being constant (see Fig. 32). 3SKE (1) is shown. The starting point of the scan is indicated by nontransparent model, this is also true for figures of corresponding scans below.

A notable relation of interaction energy and angular change could be traced in 3SKE (2) with the interaction energy maximum and minimum difference of about -3 kcal/mol. The trend is slightly less apparent in 4IBK (1) and 3SKE (3), the latter two display the general shape expected for chalcogen bond driven interaction but the dependence is more shallow (about -2 and -1.5 kcal/mol, respectively). Other systems do not display any significant dependence that would be expected for chalcogen bond featuring systems nor any other clearly apparent tendencies (Fig. 32).

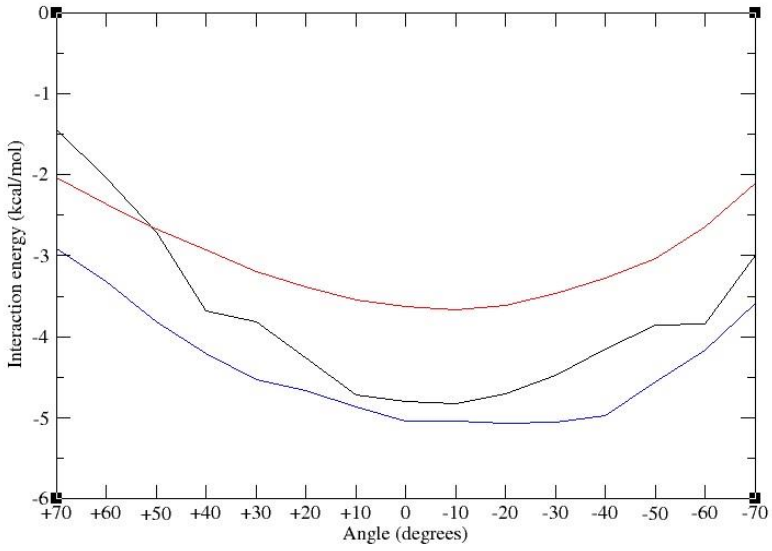


Figure 32. Angular scan 1, the relation of the angular deviation and the interaction energy

Angular change is in degrees (x axis), interaction energies are in kcal/mol (y axis). The angle of 0° denotes the starting orientation. From that the systems were shifted up to +/- 70° to each side. The minimum of the interaction energy is at or very close to the starting point, from there the interaction energy grows, which is most likely due to leaving the preferred chalcogen bond geometry. But also, since hydrogen bond is also directional, some of hydrogen bond contribution is to be considered as well. Black 3SKE (2), blue 4IBK (1), red 3SKE (3). The curve was not informative in case of other model systems.

5.3.1.6.2 Angular scan 2

(In the plane of the ring separating the hydrogen bond from the chalcogen bond)

In this angular scan, the amino acid model was turning around the nearby hydrogen atom, removing the influence of the chalcogen bond. The resulting interaction energy change should mainly be due to the hydrogen bond strength change and its angular preference. The hydrogen to oxygen distance was kept fixed (Fig. 33).

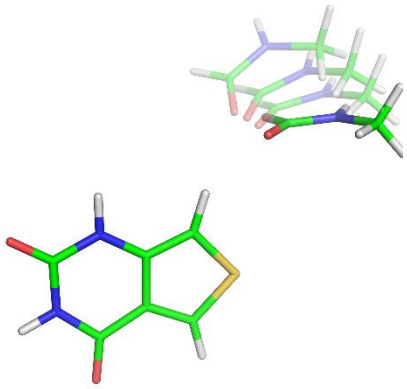


Figure 33. Scan 2, the relation of the angular deviation and the interaction energy

The scan 2 was performed with the amino acid model turning around the ligand´s hydrogen atom. This way possible contribution of the chalcogen bond is ablated while the hydrogen bond is maintained (or strengthened). The oxygen to hydrogen are kept constant (see Fig. 36). 3SKE (1) is shown.

The curves of the relation of the interaction energy to the angular change are flat and hard to interpret. 3SKE (2) and 4IBK (1) shows similar relations seen for the substituted systems (see section 5.3.2.6.2). For these two systems the interaction energy increases (becomes less negative) when moving from the start. The increase goes until the energy maximum is reached and then a decrease is apparent in Fig. 34. The interaction is weakened by 2 about kcal/mol for these two systems. The hydrogen bond should be getting stronger when C-H…O is approaching linearity, thus the chalcogen bond should have about -2 kcal/mol or more negative contribution in these systems. This relation does not occur in other systems, though.

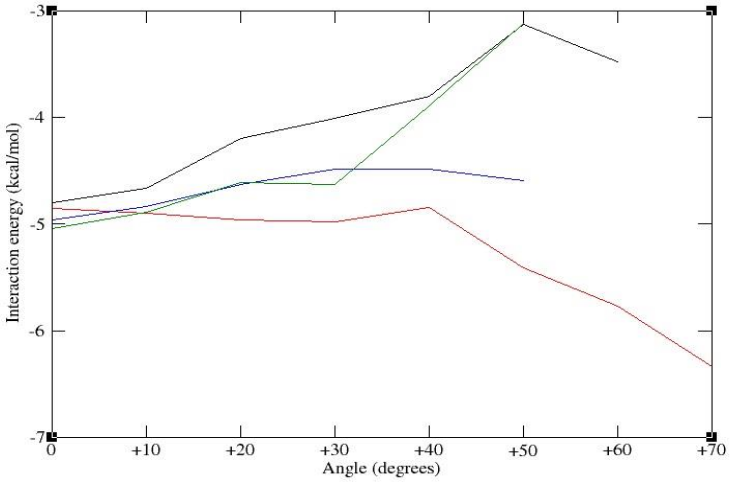


Figure 34. Scan 2, relation of the angular deviation and interaction energy.

Starting orientation is denoted by the angle of 0°. The spike in the interaction energy (at about +50° for 4IBK (1) and 3SKE (2)) probably corresponds to a point when the chalcogen bond is the weakest and hydrogen bond is not yet close enough to compensate for the loss of the chalcogen bond. The relation was most significant for 3SKE (2). Angles are in degrees, interaction energy in kcal/mol. 3SKE (2) black line, 4CTK (1) blue line, 3SKE (1) red line and 4IBK (1) green line. Systems that do not exhibit a predictive relation are not included in the graph (only 4CTK, blue line, is shown as an example of one).

5.3.1.6.3 Angular scan 3

(In the plane of the ring from one sulfur atom σ-hole to the other one)

Last angular scans were performed with the distance of sulfur to the nucleophile fixed again. Now the amino acid models turned around the sulfur atom in the plane of the ring (Fig. 35). By this motion, the effect of the hydrogen bond and the chalcogen bond are both filtered out. For the interaction energy with the C-S…O angle of about 135° (around +45° in the Fig. 35, considering the starting C-S...O angle of 180°) (where the σ-holes and hydrogen atoms are most distant from the nucleophile) nonspecific dispersion should be the major remaining component of the binding.

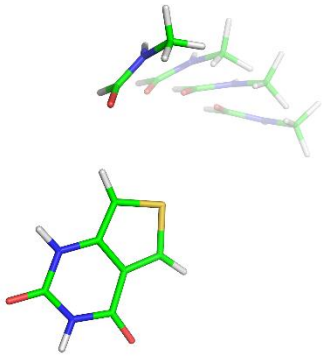


Figure 35. Scan 3, in the plane of the ring from one σ-hole to the other one (noninteracting one)

The scan 3 was performed with the amino acid model turning around the sulfur atom with sulfur to oxygen atoms distance kept constant (the turning is in opposite direction than in scan 2) (see Fig. 36). 3SKE (1) is shown.

There are maxima of the interaction energy roughly in the position where the distance of the moving oxygen atom to both σ-holes is the biggest (where no chalcogen but also no hydrogen bonds should participate significantly) (Fig. 36). In some cases (4IBK (1) and 4CTK (1)) there are maxima beyond these positions of the second σ-holes orientation (the one that is not interacting originally). This is most likely due to the steric repulsion of molecules as the starting orientation for all the systems is not the same. A turn of the interaction partner by a certain angle may cause a steric repulsion in one system, while it does not in another system. Minima of the interaction energy were found at the starting and ending angles of the scan. Even despite that in 3SKE (2) case, the sulfur to oxygen atom distance was well below the sum of van der Waals radii, the interaction energy remains negative in the course of the entire scan. The absence of repulsion for (at least) 3SKE (2) case in entire course of this scan is peculiar.

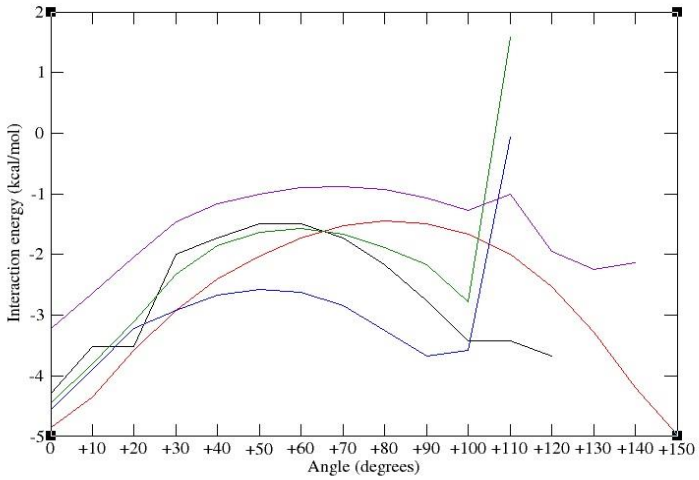


Figure 36. Scan 3, relation of the angular deviation and the interaction energy

Angles are in degrees (x axis), interaction energy is in kcal/mol (y axis). The scan starts at 0. The maximum of energy is roughly around +45°, orientation most distant from both σ-hole and hydrogen bonds. 3SKE (2) black, 3SKE (1) red, 3SKE (3) magenta, 4IBK (1) green, 4CTK (1) blue. The sharp significant interaction energy increase for 4IBK (1) and 4CTK (1) at the right corresponds to a steric clash of compounds and subsequent repulsion. System 3SKE (4) did not show predictable relation and is not shown.

5.3.1.6.4 Distance scan

The distance scan simply moved the interacting partners apart. The interaction energy was calculated each step by 10% of the sum of van der Waals radii. 10 steps were used to move the compounds farther apart and 2 steps brought the compounds closer from the starting point.

The plots generally have the same shape in all the cases (Fig. 37), which was reminiscent of that seen in Fig. 3. The interaction energy was the most negative for the starting system geometries (geometries after the second free optimization, section 5.3.1.5). It went sharply upwards (towards more positive interaction energy-repulsion) when bringing the nucleophile and the chalcogen closer or went upwards slightly less sharply when the interacting partners were moved apart.

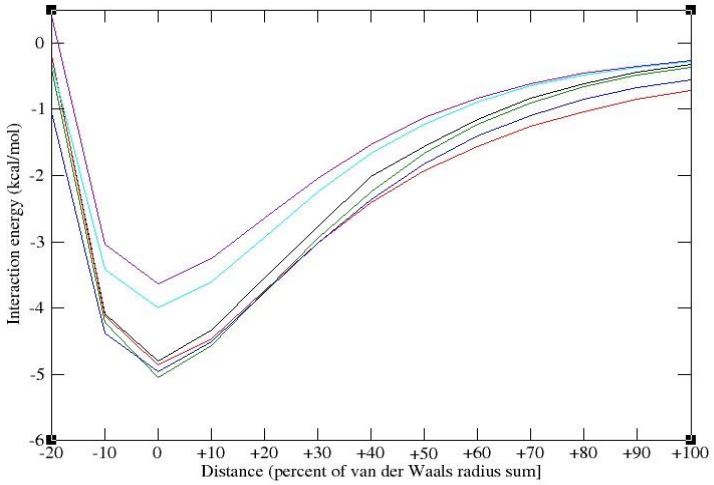


Figure 39. Distance scan

The plot depicts the intermolecular separation of the interacting compounds. Distance is in steps by 10 percent of the van der Waals radii sum, interaction energy is in kcal/mol. The interacting compounds were moved with the C-S…O angle being kept constant. The distance 0 denotes the starting orientation, from there the systems were moved apart (positive values) or brought nearer (negative values). The dependence is similar for all of the systems and closely resembles the one shown in Fig. 3. 3SKE (2) black, 3SKE (1) red, 3SKE (3) magenta, 3SKE (4) light blue, 4IBK (1) green, 4CTK (1) blue.

5.3.2 Model systems tuned for stronger chalcogen bond

Substituted systems are systems analogous to the crystal structure models but engineered to display the greater σ-hole. This is done by carbon bound hydrogen to fluorine atoms and sulfur to selenium substitutions. Electronegative fluorine atoms withdraw electronic density from the chalcogen, enlarging the σ-holes, and the sulfur replacement with a more polarizable selenium atom should also work for the greater σ-hole (as described in the section 2. Literature Review). These systems are models for optimization of chalcogen bonding properties of ligands. However in actual compound design for each system separately, unique optimization procedure should be chosen to increase the interaction strength (for instance not replacing certain hydrogen atoms relevant for the binding). Substituted ligand models described here serve only as a lead for further optimization.

The same procedure as for model systems described in the above sections was performed with substituted model systems, too, e.g. optimization and scan steps.

5.2.2.1 Electrostatic potentials

In the substituted ligand models, the σ-holes are well apparent in all the cases. This is due to the fact, that hydrogen atoms adjacent to the chalcogen atoms were replaced with fluorine atoms, thus their positive ESP do not blend with that of the σ-holes, but they rather form enclosed regions - real holes. 4U7Q and 3SKE substituted ligands have comparable and greatest σ-holes among the systems and 4CTK (A) has the smallest σ-hole(s) of all the substituted compounds. In case of substituted systems, the σ-holes were also slightly farther away from the C_8-Se_{11} axis (Fig. 38, 3SKE (A) as an example, the deviation is not apparent in this picture).

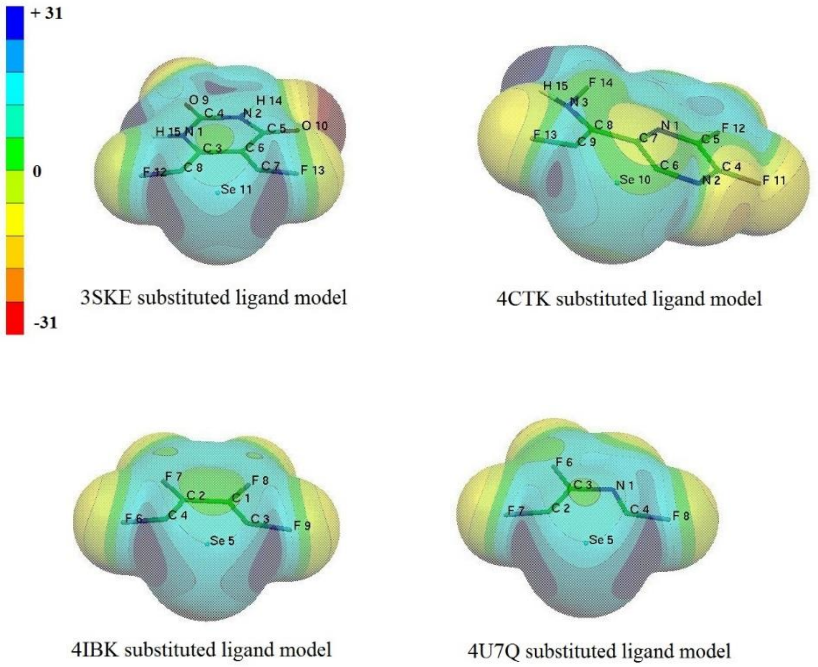


Figure 38. ESPs of substituted ligand models

The same color scale of ESP from +31 to -31 kcal/mol is used for all the systems. σ-holes are apparent roughly in the direction of the both C-Se bonds extensions as distinct bordered regions. The fluorine atoms are partially negative, thus they may be able to attract for instance hydrogen atoms via electrostatics.

Table 10. Magnitudes of ESP of substituted ligand models

Values of electrostatic potential are in kcal/mol. Interacting site is considered the one of two σ-holes that is in the concrete model system closest to the interacting partner. Magnitudes are measured at centers of circular areas of positive ESP on the sulfur atom (not exactly in the axis of the C-S bond). For symmetric molecules interacting and noninteracting sites are considered being the same and minor deviations in the measurement for both sites (not listed) are attributed to the experimental error. 4U7Q and 3SKE substituted model ligand have the greatest σ-hole among the systems. 4CTK (A) has not the noninteracting σ-hole positive (compared to 4CTK (1)), but this system still has the smallest σ-hole among the substituted systems.

Models	Magnitude (kcal/mol) Interacting site	Magnitude (kcal/mol) Noninteracting site
3SKE(A)	39.91	38.31
3SKE(B)	38.31	39.91
4CTK(A)	29.59	23.28
4IBK(A)	37.26	37.26
4U7Q(A,B)	39.52	39.89

5.3.2.2 Geometries

Initial (crystal structure) geometries for substituted systems were understandably similar to those of the unsubstituted systems. Only necessary optimization steps in preparing the substituted model systems did change certain geometrical parameters slightly. For instance, the interatomic distances have been modulated to correspond to the same percentage of the sum of van der Waals radii as in the unsubstituted systems (the sum of van der Waals radii for selenium and oxygen atoms is 3.42 Å, for sulfur and oxygen atoms 3.32 Å). 4CTK (A) and 4U7Q (B), these systems both have their crystal structure geometries close to the optimum for a strong chalcogen bond and also have the highest interaction energies among these model systems (Table 11, optimization of fluorine and hydrogen atoms).

Correspondingly to the hydrogen optimization step above (section 5.3.1.2), for this set of model systems, hydrogens and added fluorine atoms were optimized first. The most negative interaction energy was found for the 4U7Q (B) system (about -4 kcal/mol). This was also the system with the C-Se…O angle closest to linearity. The second strongest interaction energy was observed for 4CTK

(A), that also had its geometrical parameters reasonable for a strong chalcogen bond contribution. While the 4U7Q (1) had a positive interaction energy (about +9 kcal/mol), the substituted version 4U7Q (A) had only slightly positive interaction energy at the same intermolecular separation. The second weakest interaction energy belonged to 3SKE (B) that had least chalcogen bond permissive orientation (the longest Se…O distance and smallest C-Se…O angle).

Table 11. Orientations and interaction energies of substituted systems before and after the optimization

Values of interaction energy (ΔE) is in kcal/mol, distances are in ångströms, angles are in degrees.

	Optimization of fluorine and hydrogen atoms			Free optimization		
System	Se...O distance (Å)	C-Se…O angle (°)	ΔE (kcal/mol)	Se...O distance (Å)	C-Se…O angle (°)	ΔE (kcal/mol)
3SKE (A)	3.55	142°	-2.62	2.87	172°	-5.09
3SKE (B)	3.87	129°	-1.83	3.57	115°	-2.87
4CTK (A)	3.24	163°	-3.40	3.04	173°	-3.87
4IBK (A)	3.64	160°	-2.79	2.83	173°	-6.41
4U7Q (A)	2.56	142°	+0.46	2.93	169°	-5.50
4U7Q (B)	3.32	177°	-4.20	2.86	165°	-5.45

5.3.2.3 Free optimization

After the optimization all of the substituted systems except for one adopted the geometry eligible to a chalcogen bond which corresponded to the interaction energies becoming more negative (Fig. 39, Table 11). With the exception of 3SKE (B) every system adopted the C-Se...O angle of about 170°. The interaction energies were slightly increased with the respect to not substituted systems overall (Table 11, Table 7). The least negative interaction energy (-1.83 kcal/mol) belonged to 3SKE (B), that did clearly not have the optimal chalcogen bond geometry. The most negative interaction energy (-6.41 kcal/mol) was the one of 4IBK (A) substituted with most fluorine atoms out of this series. The interatomic distance seems to correlate with the interaction energy to a certain degree (the most negative for the lowest interatomic distance, the least negative for the greatest

distance). Only 4CTK (A) had the interaction energy lower than its unsubstituted variant - 4CTK (1), while also having a chalcogen bond permitting geometry (-3.83 kcal/mol for substituted and -4.74 kcal/mol for not substituted system).

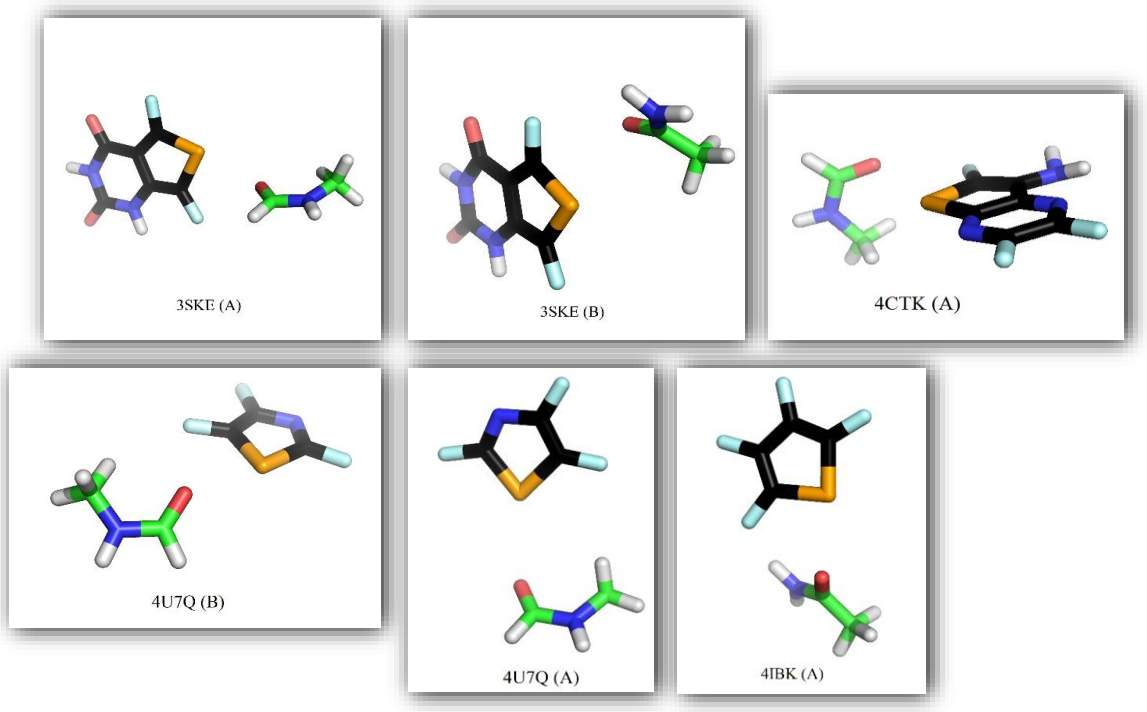


Figure. 39. Substituted model systems after the free optimization step

Carbon atoms are green for amino acid models, black for ligand models, oxygen atoms are red, nitrogen blue, selenium orange, fluorine light blue and hydrogen atoms are white.

5.3.2.4 Presumed optimal chalcogen bond geometry optimization

Forcing the substituted systems C-Se...O angle to 180° did not in fact make the interaction stronger, but there was a slight decrease in the interaction energy strengths instead (interaction energies became less negative) in all the systems except for 3SKE (B). 3SKE (B) was in an orientation unsuitable for a chalcogen bond before (see Table 12). 4IBK (A) after this step was still the one having the strongest interaction (-5.70 kcal/mol) and 4CTK (A) among the weakest (-3.45 kcal/mol).

Table 12. Substituted model systems after the optimization to the presumed chalcogen bond optimal geometry

Interaction energies (ΔE) are in kcal/mol, distances are in ångströms.

System	Se...O distance (Å)	ΔE (kcal/mol)
3SKE (A)	3.38	-4.00
3SKE (B)	3.15	-3.45
4CTK (A)	2.92	-3.45
4IBK (A)	2.96	-5.70
4U7Q (A)	3.15	-3.88
4U7Q (B)	2.98	-4.67

5.3.2.5 Free optimization from the chalcogen bond optimal geometry

In this step model systems adopted roughly the same geometry (see Table 13) they had before the restraint optimization to C-Se…O linearity (after the optimization step, section 5.3.2.3). Only the 3SKE (B) system that formed the chalcogen bond in previous step has not returned to the) obviously energetically suboptimal (in this case) state. The interaction energies and Se…O distances are also very comparable to those after the 5.3.2.3 step. 4CTK (A) has the weakest interaction and 4IBK (A) the strongest one. The C-Se…O angle did not remain at the 180° forced in the previous step for neither system (the angles are between 172° and 165° after this step).

Table 13. The second free optimization of the substituted systems

The geometry criteria and interaction energies (ΔE) of the substituted model systems after the second free optimization (following the optimization to presumed chalcogen bond optimal geometry). Energies are in kcal/mol, distances are in ångströms, angles are in degrees.

	Se...O distance (Å)	C-Se...O angle (°)	ΔE (kcal/mol)
3SKE (A)	2.85	172°	-5.79
3SKE (B)	2.90	170°	-4.65
4CTK (A)	3.09	170°	-4.03

4IBK (A)	2.85	172°	-6.40
4U7Q (A)	2.92	170°	-4.26
4U7Q (B)	2.88	165°	-5.19

5.3.2.6 Scans

Angular and distance scans were done correspondingly to section 5.3.1.6 - like in unsubstituted system scans. Only instead the hydrogen atoms, fluorine atoms were considered, even despite of their contribution to the interaction energy is presumably much smaller.

5.3.2.6.1 Angular scan 1

In the substituted systems the motion of the oxygen atom away from the ring plane resulted in interaction energies with a quadratic-like curve that had the minimum in the initial geometry (Fig. 40). 4CTK (A) exhibiting the least negative interaction energy also shows the weakest angular dependence of all substituted systems - about 2 kcal/mol difference between the maximum and the minimum. The difference is the greatest for 4IBK (A) (about 5.5 kcal/mol). It seems that the interaction energy correlates with the directional tendency.

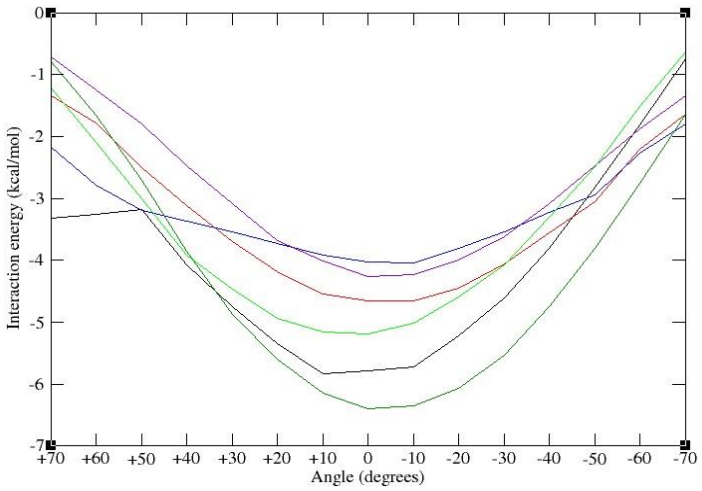


Figure 40. Scan 1 of the substituted systems, the relation of the angular deviation and the interaction energy

The angular change is in degrees (x axis), interaction energies are in kcal/mol (y axis). The 0° point denotes the starting orientation. From there systems are shifted up to +/- 70 degrees to each side. The minimum of interaction energy is at or very close to a starting point. From there the interaction energy grows (becomes less negative), which is most likely

due to the leaving of preferred chalcogen bond geometry. Unlike in Fig. 32, there is no hydrogen bond except for possible weak ones in some cases. 3SKE (A) black, 3SKE (B) red, 4IBK (A) green, 4CTK (A) blue. 4U7Q (B) light green, 4U7Q (A) magenta.

5.3.2.6.2 Angular scan 2

In case of the substituted systems, the interaction energies exhibited curves with the minimum at the starting point and the energy plateau from roughly +40° onward (Fig. 41). This shape of a curve was very similar in all the studied systems. There is no spike of interaction energy like there was for some unsubstituted systems displaying similar relation (5.2.1.6.2). The difference of interaction energies before and after the scan was from about 3 to 4 kcal/mol, suggesting that this is the contribution of the chalcogen bond.

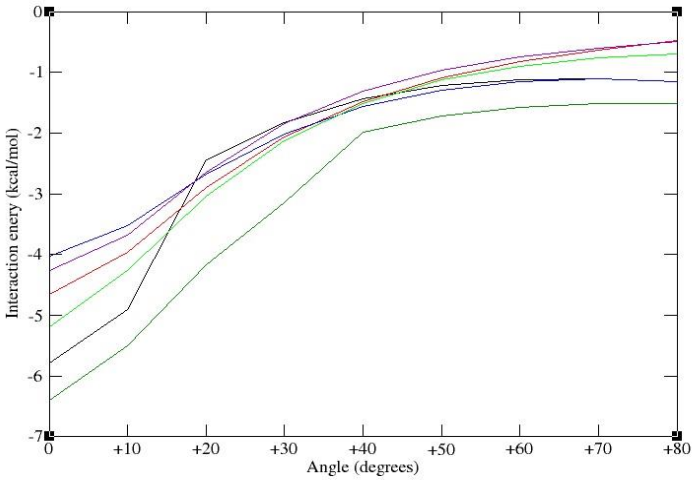


Figure 41. Scan 2 of the substituted systems, relation of the angular deviation and the interaction energy.

Starting orientation is denoted by point 0 on the x axis, from there scan up to +80° is performed. The relation was similar for all the substituted systems. There is an interaction energy increase due to the loss or weakening of the chalcogen bond followed by an interaction energy plateau. Negative interaction energies for systems in the plateau are probably due to the nonspecific dispersion or weak hydrogen bond becoming stronger. 3SKE (A) black, 3SKE (B) red, 4IBK (A) green, 4CTK (A) blue, 4U7Q (B) light green and 4U7Q (A) magenta. Interaction energy on y axis (kcal/mol), angular shift from original position on x axis (degrees).

5.3.2.6.3 Angular scan 3

The hill-like curve (A-shaped) was plotted by interaction energies of substituted systems where amino acid model shifted from one σ-hole to another. The maximum never exceeded -2 kcal/mol (Fig. 42), unlike in original unsubstituted systems (Fig. 36).

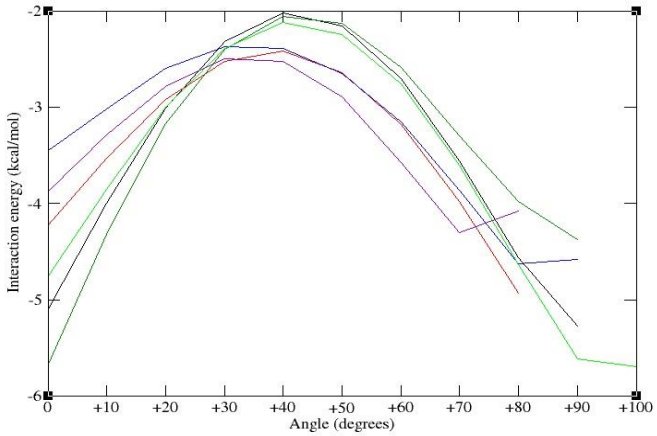


Figure 42. Scan 3 of substituted systems, the relation of the angular deviation and the interaction energy

Interaction energy on y axis (kcal/mol), angular shift from original position on x axis (degrees). As for unsubstituted systems the energy maximum is about in +45° from the starting orientation (most distant from both σ-holes. 3SKE (A) black, 3SKE (B) red, 4IBK (A) green, 4CTK (A) blue, 4U7Q (B) light green, and 4U7Q (A) magenta.

5.3.2.6.4 Distance scan

Separating molecules apart gave similar results as for unsubstituted systems (Fig. 43). The only notable difference between substituted and unsubstituted system was that for substituted systems did the interaction energy rocketed more when the Se...O distance was shortened compared to the unsubstituted systems.

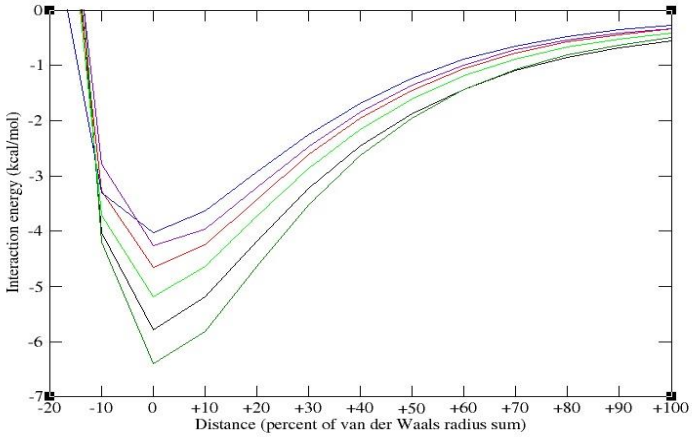


Figure 43. Distance scan of substituted systems

Plot corresponding to that of fig 38, but for the substituted systems. The interaction energy on y axis (kcal/mol), interatomic selenium to oxygen distances are in the percentage of the sum of van der Waals radii, on the x axis. 3SKE (A) black, 3SKE (B) red, 4IBK (A) green, 4CTK (A) blue, 4U7Q (B) light green and 4U7Q (A) magenta.

6 Discussion

6.1 The amount of motifs discovered

Seemingly only few proteins-ligand complexes with fitting parameters were found in the PDB database. From 217 structures, 4 met the criteria for further examination, which is 1.84 % of structures. However, this small number is comparable to the number of close contacts (from 0.7 to 18%) discovered by Bauzá et al. (2013) in their Cambridge Structural Database search for various σ-hole interaction motifs of. On the other hand, Lu et al. (2009), who searched the PDB database for halogen to oxygen close contacts found out that about one sixth of the searched compounds fitted their criteria (distance below the sum of van der Waals radii and the C-X…O angle not below 140°). Considering viral protein restriction, the number of found close contacts is not that negligible as it seems. The number of the chosen structures is still too small for proper statistical analysis, though. Sufur is not among the most abundant elements in protein ligands, divalent sulfur even less so. This may also be reflected in this low absolute number of discovered close contact, and should be noted before any considerations regarding properties of interactions are done.

6.2 Electrostatic potentials

Values of magnitudes of electrostatic potential overall are substantially greater than what is stated in the literature for σ-hole exhibiting chalcogen atoms (Wang et al., 2009). Strongly positive ESP around hydrogen atoms that are in the close proximity to chalcogens may somehow influence the magnitude. The fact that σ-holes were the most prominent in 3SKE (1) and (2) systems and insignificant in 3SKE (3), (4) and 4IBK (2) (thiophene ring) was in accord with the theoretical expectations. The tiophene ring lacks any electronegative, electron-withdrawing, groups. From the comparison of ESP it seems, that the chalcogen bond (σ-hole) is more tunable than the hydrogen bond (partially positive hydrogen atom). Thiophene ring has only insignificant extensions of positive ESP from hydrogen towards where the σ-hole should be. As opposed to this, 3SKE (1) and (2), more closely resembling the original ligand tuned with a condensed uracil ring, had very notable σ-holes, one of them even being more positive than the nearby hydrogen atom - 3SKE (2).

The noninteracting σ-hole of 4CTK (1) was of the least positive of the studied systems. The reason for such asymmetry of two σ-holes on a single compound may possibly be due to the local environment, e.g. proximity of the partially negative nitrogen atom. The asymmetry of 4IBK (1) σ-holes could also be ascribed to the local environment; there is a chlorine atom possibly interfering with a significant σ-hole formation site. While fluorine atoms facilitate σ-hole enlargement, *proximal* chlorine or nitrogen atoms (even despite of being more electronegative than heavier chalcogen on carbon atoms) may work against the strong σ-hole. This effect can be simply explained by the combination of partial positive (electron deficient) regions on the σ-hole with partial negative regions (electron rich) on the proximal compound. The adjacent chlorine atom is bigger and less electron withdrawing, thus it may diminish the proximal σ-hole to a certain degree. Another effect diminishing the σ-holes of 4CTK compounds is probably their weakening by NH_2 substituent that is an electron-donating group (Murray et al., 2008).

The systems modified with fluorine and selenium substitutions had all significant σ-holes. 4CTK (A) has the smallest σ-hole of all the substituted compounds, presumably because only one fluorine atom is adjacent to the selenium, other substituted fluorine atoms are on the other side of a compound (additionally to reasons noted above for 4CTK (1)). Interestingly, the fluorine atoms themselves are only mildly negative – much less than an oxygen atom despite of their greater electronegativity. It seems, that the electronegativity alone is not a sufficient parameter to predict the ESP (or the partial charge), as was suggested by Murray et al. (2014).

6.3 Geometries and interaction energies in the experimental setup

6.3.1 Crystal structure geometries

Most chalcogen bond suitable starting (derived from the crystal structure) geometries are exhibited by model systems 4CTK (1) and 4U7Q (2). 4CTK (1) has a reasonably negative interaction energy. The interaction energy of 4U7Q (2) is however less significant (see Table 7). 4U7Q (2) is very close to the theoretical optimal chalcogen bond angle, but its angular deviation from theoretical C-S…O angle of 180° is to the opposite side as in 4CTK (1), that is over the edge of a C-S-C (away from the hydrogen atom and also away from the actual σ-hole – see below). The interaction energy is also much less negative for 4U7Q (2) than for 4CTK (A) after this step. 4U7Q (1) has a very large and positive interaction energy. This suggest possible error in the original crystal structure or that this orientation is forced by other interacting moieties. Since the protein is a symmetric dimer

there is a little reasoning that such close S to O distance in 4U7Q (1) would be forced by other interacting moieties, because the orientation is much more reasonable in 4U7Q (2), which is an alternative binding mode for a ligand in this *symmetric* protein dimer. The flexibility of the viral protein is something to take into account, too.

Among the systems substituted by fluorine and selenium atoms, the two systems with the most reasonable chalcogen bond geometry, 4CTK (A) and 4U7Q (B), had the greatest (the most negative) interaction energies. Only 4CTK (A) had the interaction energy smaller than its unsubstituted variant 4CTK (1). This is possibly due to the absence of the hydrogen bond in the substituted variant 4CTK (A) and due to the lack of a chalcogen bond of greater strength (for reasons noted above) - the chalcogen bond did not cover the energy lost by the hydrogen bond removal. 4U7Q (A) had almost 9 kcal/mol more negative interaction energy than its non substituted variant 4U7Q (1), even though their geometries are analogous. This extreme difference is observed even despite of 4U7Q (A) before free optimization is nowhere close to chalcogen bond permitting geometry. Seemingly selenium atom in substituted systems is capable of significant nonspecific dispersion driven interactions.

6.3.2 Free optimization

In the free optimization most of the model systems got into orientations less favorable for a chalcogen bond in terms of the sulfur to oxygen atoms distance and the C-S...O angle. Yet there was an apparent decrease in the interaction energy. This decrease is most likely due to the secondary interaction (hydrogen bond) formation, since the hydrogen atoms of the ligand models got closer to the amino acid model oxygen. However, from the fact, that systems did not optimize straight to the C-H...O linearity (that would be presumed to be the most optimal for a hydrogen bond), the chalcogen bond may still be at play – a competition of both interactions is probably the case. Such competition between hydrogen bond and σ-hole interaction has been described in the literature (Syzgantseva et al., 2013). In 4U7Q (1) the system relieved an obvious steric stress by moving the interacting partners apart, as it is apparent from the significant drop in the interaction energy. Since there is no secondary interaction for the geometry this model system adopted after free optimization, and the chalcogen bond is also not probable to contribute due to the inappropriate interacting partners orientation, most of its interaction energy (-1.07 kcal/mol) is to be ascribed to the (nonspecific) dispersion between sulfur and oxygen atoms. There is also a slight trend of the

hydrogen bond getting stronger with electron-withdrawing residues, judging from the interatomic distances (see 3SKE (1) and (2) compared to (3) and (4); 4IBK (1) compared to (2), Table 7). When optimizing the compounds for a chalcogen bond by addition of electronegative or electron-withdrawing substituents, any hydrogen bonds that a compound can form may also get tuned by a certain degree. Even though in 4CTK and 4U7Q the interacting partners were in the chalcogen bond favoring orientation before the free optimization, this orientation is probably not solely caused by the influence of a chalcogen bond, as during optimization the corresponding systems adopt geometry where the chalcogen bond presumably plays much lesser role. This may be the effect of other interacting moieties that force the geometry in the protein crystal to what would be interpreted as a chalcogen bond. Even so, the chalcogen bond may play a lesser stabilizing role here. Since the geometry in the crystal already is close to chalcogen bond optimal geometry for these two cases, further optimization exploiting the chalcogen bond of these ligands would be less likely to encounter any difficulties caused by the ligand being unable to accommodate to the σ-hole-to-nucleophile orientation properly.

Judging from the change of interacting partners orientations of substituted systems during the free optimization procedure and from corresponding interaction energies, chalcogen bond was most likely formed or strengthened in this process for all the systems but 3SKE (B). This is probably why substituted system 3SKE (B) has the least negative interaction energy in the free optimization step. As there theoretically should not be other strong noncovalent interactions in the substituted systems (the strong nucleophilic oxygen is lacking positive residues on the ligand to interact with, as it is apparent from ESPs (Fig. 38), most of 3SKE (B)´s interaction energy after this optimization step should be ascribed to dispersion, giving the approximate value of "non-σ-hole interaction energy" for this system or substituted systems in general. Obviously when the systems accommodate towards the chalcogen bond, relative ratios of interaction energy contributions from the system´s functional groups change – it is hard to make any generalizations or predictions of the interaction energy contributions for various functional groups.

The second least negative interaction energy has been observed for 4CTK (A) as it is increasingly apparent that the chemical environment of the selenium atom of this 4CTK ligand models attenuate the σ-hole to a certain degree. 4IBK (A), tetrafluorothiazole, has the greatest interaction energy among the system and greater than trifluoroselenazole rings of 4U7Q (A) and (B) systems. It would

seem plausible, that one additional fluorine atom in 4IBK (A) is responsible for a greater electron withdrawal than the nitrogen atom as heteroatom in 4U7Q substituted models. However from comparison of ESPs it is trifluoroselenazole (4U7Q) that has slightly greater σ-hole. If the σ-hole are considered for both ligand model comparable a simple subtraction would give the point of reference value of the interaction energy difference between these two systems given by secondary interaction. The secondary interactions (possibly F...H-N interaction of 4IBK) are about -1 kcal/mol strong. This is under the assumptions that other than σ-hole interactions are not the major contribution on the interaction energy of substituted systems. ESPs of substituted systems depicted position of σ-holes consistently with the C-Se...O angle observed after the optimization – deviating from the theoretical position.

6.3.3 Optimization into presumed chalcogen bond optimal geometry

Interestingly, after the 180° C-S...O geometry has been forced to the model systems (that is according to results around 10° off of the actual optimum) some of them attained or maintained the sulfur to oxygen atom distance below the sum of van der Waals radii. It is reasonable to presume that these systems, 3CKE (1) and (2); 4CTK (1); 4U7Q (2), have some chalcogen bond component in their interaction energy. These are among the systems that did optimize to the geometry in which the chalcogen bond played at least partial role, except for the 4U7Q (2). In the 4U7Q (2) system despite of some chalcogen bond contribution might be present, the hydrogen bond is probably much stronger as its free optimization suggested (see section 5.3.1.3). The interaction energy was in case of every system except for 4U7Q (1) less negative than in previous step. This interaction energy increase most likely corresponds to the weakening of the hydrogen bonds of respective systems. However it is not possible to completely distinguish the hydrogen bond contribution, as now the hydrogen bond probably still participates to a certain unknown extend. Concerning the 4U7Q (1) system in the previous step (section 5.3.1.3) and in this 180° forced optimization alike, there were no significant interactions other than dispersion. Thereby now the 4U7Q (1) energy drop is mostly because of the chalcogen bond formation it is however very modest drop of the interaction energy, of about -0.8 kcal/mol.

6.3.4 Free optimization from the chalcogen bond optimal geometry

Altogether, most of the crystal derived model systems were not in the geometry suitable for a chalcogen bond after this step. The interaction energy decrease due to the strengthening of the hydrogen bond outweighs the interaction energy increase (decreasing negativity) from leaving the optimal chalcogen bond geometry as it seems. In some systems chalcogen bond was not completely excluded, however most likely (and in most cases) the hydrogen bond is stronger (as was discussed in above sections). After the free optimization from the presumed chalcogen bond optimal geometry, 3SKE (2) returned to approximately same position as it was before, confirming the energy minimum located in the first optimization step.

Out of 4IBK systems, (2) - without a chlorine moiety - optimized roughly to the geometry with an oxygen atom facing the C-H bond. On the other hand, 4IBK (1) with the chlorine might even after this step have some chalcogen bond contribution. This might mean that the chlorine moiety of 4IBK (1) facilitated the greater σ-hole, which subsequently influenced more strongly the optimization process than in 4IBK (2), since both systems have otherwise similar geometry. However, the same system as 4IBK (2) only with a different starting orientation - 3SKE (3) did not shift towards the hydrogen bond formation entirely while having smaller interaction energy than 4IBK (2) that did shift. The software most likely did not locate the same minima for different starting geometries and neither system actually has a strong chalcogen bond. The latter can imply that only one step can generate false results (with calculation at used level of precision).

Concerning substituted systems, after this step most of them returned to where they were before the C-Se...O 180° forcing optimization. This confirms where energy minima connected with chalcogen bond formation really are. 3SKE (B) which deployed the chalcogen bond orientation in previous step maintained it after the second optimization, showing the true local minimum and failure of a software to find it after just first optimization. From the geometries of the substituted systems after the free optimization it seems than when the chalcogen bond is dominant the C-S...O angle does not drop too much below approximately 165° (see Table 13.). In contrast to this, the systems not substituted for the strong chalcogen bond always had this angle below 165° (see Table

9). Only the 3SKE (2) had the C-S...O angle 165° (162° after the first free optimization) - the difference in C-S/Se...O angles of 3SKE (2) and substituted systems is small. This system also exhibits more positive σ-hole (even though the respective ESPs are comparable) than nearby hydrogen atom – unlike other unsubstituted systems. Thereby it seems that if chalcogen bond is dominant, the position of nucleophile does not deviate too much from where the ESP shows the σ-hole in spite of secondary interactions. Following that logic, in other unsubstituted systems chalcogen bonds probably are minor contributors, as their ESPs suggest.

3SKE (2) consistently retains characteristics expected form a system in which chalcogen bond is a dominant factor. However in the original protein ligand crystal, this model does not match an interaction motif that is prominent in the ligand´s binding. Slight accommodation of the chalcogen bond optimized ligand would be necessary for the stronger binding. On the other hand ligand models 4CTK (1) and 4U7Q (2) had their starting orientations much more suitable for the chalcogen bond, but the chalcogen bond did not prove to be the driving force for a binding. The substituted models, with fluorine and selenium substitutions, on the other hand were capable of the strong chalcogen bond formation. Thus 4CTK and 4U7Q ligands, which are in the orientation suitable for the chalcogen bond formation or strengthening, can be modified the way to exploit this interaction. For example by replacing the sulfur atom by heavier chalcogen selenium or tellurium or by implementation of electron withdrawing groups.

Overall, despite the fact there were signs of a weak chalcogen bond apparent in so far discussed experimental steps in some of model systems, it does not appear that chalcogen bond is the dominant factor for a binding in studied crystal structures.

6.3.5 Scans

For scan 1 substituted systems provided all well-apparent U-shaped curve where interaction energy maxima correspond to the chalcogen bond loss. From those, 4CTK (A)´s relation of interaction energy and angle was most shallow. Only unsubstituted systems with similar relation were 3SKE (2) and to lesser extent 4IBK (1) or 3SKE (3). In previous steps 4IBK (1) or 3SKE (3) did not consistently seem to be significantly better in terms of chalcogen bond contribution than other systems. Conversely, some systems that showed some of the σ-hole interacting properties before

did not exhibit this U-shaped relation. Differences in positions of other interacting groups among systems may cause these inconsistencies (for instance further steps of angular scan are energetically disfavored by a repulsion in one system). 3SKE (2) and 4IBK (1) showed some of the tendency seen in substituted systems, but others did not. However even systems that did not show much of directional preferences had relatively high negative interaction energy. From these facts it seems, that the highly negative interaction energy in these systems are mainly because of the hydrogen bond and that the hydrogen bond in this setup is much less direactional than chalcogen bond. Therefore hydrogen bond in this setup is less likely to produce false positive results.

In scan 2 the increase of interaction energy by turning the oxygen atom containing amino acid model around the hydrogen atom of model ligand goes on until it reaches maximum and then a decrease is observed. The decrease is probably corresponding to the formation of a second hydrogen bond. The spike or energy maximum, corresponding most likely to the point where the significant portion of chalcogen bond contribution is lost, but is not fully compensated by second hydrogen bond creation is most apparent in 3SKE (2) and also 4IBK (1). This relation is noteless in other systems, though (Fig. 36). No such spikes were apparent in substituted systems, where no secondary interactions are considered. The negative interaction energy in a plateau (see Fig. 42) is probably due to nonspecific dispersion.

Scan 3 showed similar relation for substituted and unsubstituted systems. This is expectable because substituted and unsubstituted systems alike have moieties capable of interacting with nucleophile at corresponding locations. From the interaction energy maximum it is also apparent that the selenium atoms of substituted systems are capable of greater dispersion overall than the sulfur atoms of unsubstituted systems. Other angular scans for substituted systems did show the effect of dispersion (on other atoms mostly) to be slightly smaller since after the breakage of chalcogen bond the interaction energies were not as negative. Dispersion on the selenium atoms is thus stronger than on other possibly interacting moieties in these scans. It also seems, that searching sulfur or selenium atom to nucleophile close contacts based on their van der Waals radii may be misleading. Se...O distances well below the sum of van der Waals radii (since this value is not changing in the course of the scan) in substituted systems together with the negative values of interaction energy were observed even away from the σ-hole. Therefore no significant repulsion

between Se and a nucleophile occurs at low interatomic separation even though there is no strong contribution of chalcogen bond after a certain number of steps of scan 3.

Distance scans were similar for substituted and unsubstituted systems alike. The only noticeable difference is that in substituted systems the interaction energy rocketed more sharply up when bringing the interacting partners together. Probably in the substituted systems the interacting partners are already too close and further reduction of their separation in energetically not favored. This possibly maps the electron-density space around the sulfur and selenium in their respective molecules. The general shape of a curve implies only, that there is an interaction present and it does not seem possible to distinguish between interaction types. Also the optimization software did localize the energy minimum reliably since there is a sharp energy increase on both side of it (moving atoms further and closer).

The important notion is that in certain scans the interacting models indeed have different relative starting geometries. Motion of ligand model in frame of one model systems by certain amount of degrees may encounter steric hinderance while in other system the motion by same amount of degrees may not. This can cause some inconsistencies.

6.3.6 Concluding remarks on σ-hole tuning in ligand

By modifying ligand models with hydrogen to fluorine and sulfur to selenium atoms substitutions, ligand models with enhanced σ-hole bonding capacity were produced in silico. Interaction energies for these model systems were relatively higher compared to of unsubstituted systems despite of the loss of hydrogen bond contributions. This could be promising start for the optimization of certain medical compounds for a stronger chalcogen bond. In actual chalcogen bond-optimized compounds, the interacting hydrogens can be maintained not substituted), so that the positive electrostatic potential around them with that of σ-hole on a chalcogen can provide an additional stabilization of the interaction.

6.4 Comparison of chalcogen bond to halogen bond

The interaction energies obtained were of the same magnitude as those presented by Bauzá et al. (2013), who also examined systems similar to substituted systems in this work (perfluoro substituted chalcogen containing compounds). Halogen bonds found in biological systems exhibited seemingly weaker interaction – greatest of about -2 kcal/mol (Lu et al., 2009). However

it is necessary to consider the effects of secondary interactions, as hydrogen bonds are more exposed to compete with the chalcogen bond (systems studied in this work) than with halogen bond. The chalcogen bond being weaker than halogen bond may cause it to be apparently less directional. This is well apparent in Fig. 41, where the compound with the least negative interaction energy apparently has the weakest relationship between the angle deviation and the interaction energy. This may be a consequence of an electronic structure of σ-hole that could strengthen more in its maximum than in the periphery.

6.5 Model system and the original crystal

The important difference between model systems and the situation in the original crystal structure is the presence of additional residues interacting with the ligand. Thereby optimization step of isolated interacting partners may find the local energy minimum, but the question remains whether it is also the local minimum in the actual protein-ligand complex. On the other hand, fitting the molecule into an electronic density cloud when "interpreting" the results of a structure analysis may not provide a product completely authentic to the actual protein-ligand complex either. The computational analysis may provide a help or verification for such a fitting procedure (especially when all of the interacting residues are considered in calculations). For example one of 4U7Q alternative binding conformations in symmetric dimer exhibited great repulsion in out model system, while the other slight attraction. The former can probably be alleviated by a simple rotation of the thiazole ring.

6.6 Separability of interactions

As it was often the case in the (unsubstituted) model systems, chalcogen bond was not the only contribution to the binding of the interaction partners solely. Secondary interactions (hydrogen bonds), most often between the nucleophile and the ring hydrogen, were formed in the majority of unsubstituted systems in various steps of the computational procedure. It has proven not to be possible to completely separate the contributions of these secondary interactions from the chalcogen bond on the total interaction energy. During the scans that were supposed to isolate the chalcogen bond from the hydrogen bond, significant interaction energy change did not occur consistently for all unsubstituted systems. In angular scan 2, this is most likely due to the formation of a second hydrogen bond that outbalanced the energy loss from the breaking of the chalcogen bond to a certain degree. Although a certain trend was apparent in this scan possibly due to the

chalcogen bond loss in some of the systems (see Fig. 41), there was not possible to distinguish how much of the interaction energy exactly has been compensated by the formation of a new hydrogen bond (2.). Angular scan 1, which moved the nucleophile from the plane of the ring, was not distinctive enough for the unsubstituted systems either. For most cases of unsubstituted system the energy loss due to this movement was rather modest (we should also consider that hydrogen bond displays some kind of directionality too that influences the scan to an unknown degree). This modest change compared to the more significant interaction energy change (to less negative values if the interaction energy) in the substituted systems overall suggests that the binding modes of the most, perhaps except of 3SKE (2), of unsubstituted systems and substituted systems are different in terms of angular sensitivity, possibly due to the smaller contribution of chalcogen bond in the unsubstituted systems. But it is still not possible to tell what part of interaction energy can contribute to chalcogen bond alone in unsubstituted systems. Substituted systems did display more significant relations on these angular scans, thus it seems plausible that chalcogen bond plays major role in these cases (as it is understandable since no other major interaction should theoretically be present).

6.7 Deviation from the theoretical σ-hole position

It is unknown, why the σ-holes on chalcogen atoms and subsequently the formed chalcogen bonds are slightly off the theoretical value of 180° C-S/Se...O angle - the oxygen atom is slightly off the axis of the C-S/Se bond. Possible influence of ESP of the nearby hydrogens is to be excluded (at least as the only factor) as similar deviations are observed even in the case of the substituted model systems (where the ligand has no hydrogen atoms or distinct positive areas neighboring the σ-holes). The knowledge gained from both EPSs and optimization are consistent on this matter. Even though there is no strong hydrogen bond to be formed, the substituted model systems still in all the cases moved away from the theoretical orientation. In the optimization of the substituted systems (where the σ-holes are the greatest) the systems ended up with the C-Se...O angle around 170°, the angle was never close to or exceeded 180°. The ESP also shows that the σ-holes around that area. Study of σ-holes of symmetric aliphatic molecules containing sulfur (dimethyl sulfoxide) did not reveal any notable deviations from their theoretical position (Clark et al., 2008). Possible effect of the ring aromatic system on chalcogen electrons might be the cause or a contributing factor. Explaining this solely in terms of hybridization is rather difficult, because hybrid orbitals still have

their nonbonding lobes in the axis of the covalent bond. Another effect may be in play. The strain of the covalent bond of the chalcogen in the aromatic ring could possibly cause this deviation. Certain amount of *s-p* hybridization on chalcogen atom is apparent on the 96° C-Se-C angle of perfluorodimethylthioether, which is greater than 90° that would be typical for nonhybridized orbitals. In the studied ring substituted systems this angle was always smaller than in dimethylthioether as a reference molecule, for instance the 3SKE substituted ligand had the C-Se-C angle of about 85° (see Fig. 43). For the unsubstituted ligand models the situation is similar. An aromatic ring may apply a steric restraint on the geometry of chalcogen bonds. Presumably C-Se or C-S bonds are bent where their anibonding orbitals (or nonbonding – see section 2.3.3.7, subsection Ideas and conclusions) are still more or less not deformed. This would create an illusion of σ-holes being shifted from the theoretically predicted orientation in the axis of a C-S (Se) bonds (but rather the bonds themselves are shifted). For a linear chalcogen bond model the σ-hole is more aligned within the expected position (Fig. 43), but not completely - so other effect may play a role.

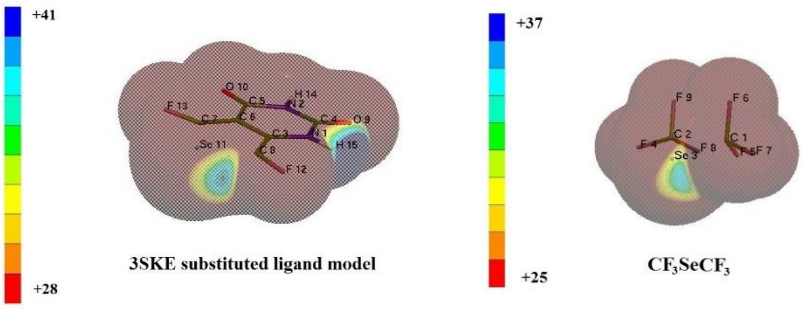


Figure 43. Position of σ-holes on a ring and a linear compound

Figure shows σ-hole position on substituted 3SKE ligand model (left) and model of aliphatic compound (all the studied compounds were rings) optimized for a significant, σ-hole perfluorodimethylselenoether correspondingly to the substituted systems (right; this system is not among studied systems). It is apparent that σ-hole is off the C-Se axis in the 3SKE model compound, where it is much closer to it in the model compound on the right. The reason why even aliphatic systems is not aligned within C-Se axis perfectly might be that the optimization procedure did not recognize the global minimum of C-Se-C angle being optimal for this system.

6.8 Future perspectives

Crystal structures of protein-ligand complexes not only of viral proteins but also of other pathogens or tumors or lower resolution crystal structures can be examined for σ-hole interactions, not only

a chalcogen bond, providing more model systems for more information and statistical analysis. Furthermore, complexes with interatomic chalcogen (or any other σ-hole exhibiting atom) to nucleophile distances greater than the sum of van der Waals radii can also be examined for the possibility to exhibit a σ-hole interactions. In experimental strategy, various other approaches can be thought of that can help to separate the σ-hole interaction from the hydrogen bond part on the total interaction energy. And with multitude of these approaches, again statistical analysis can be used to quantify better roles of both interactions. Interaction energies in a solvent can be calculated to have a better view of a biological context. Since σ-hole interactions should be less prone to the interaction energy reduction by solvation this could reveal σ-hole interactions significance where the hydrogen bond seemed to be dominating. Interaction energies, or at least some of them, can be calculated by more exact quantum-chemical methods. Energy decomposition for σ-hole interactions can be also studied by more advanced quantum mechanical methods. Lastly, ligands optimized for the stronger σ-hole interaction can be designed for concrete systems and actually synthesized and tested as potential inhibitors.

7 Conclusions

As a result of this work it can be concluded:

1) Chalcogen bonds are not an abundant interaction motif (speaking in absolute numbers) in the crystal structures of viral protein-inhibitor complexes with high resolution provided up to date.

2) The strength of the chalcogen bonds in the most of the model complexes was smaller than of the hydrogen bonds. Some of the model systems in the crystal structure derived geometry lie close to the presumed chalcogen bond-connected minimum, that is however not the local minimum of interaction energy. Chalcogen bond can help to coordinate the studied ligands by binding to the nucleophile of a protein additionally to hydrogen bonds and other noncovalent interactions, but is not the major contributor to the binding. Additionally, we have found out that for the studied systems chalcogen bonds in divalent chalcogen containing aromatic rings are situated slightly off the position where it should be according to the theory.

3) When however we optimized the model ligands for the stronger chalcogen bond, the interaction energy was in most of the cases more negative (favorable) than for systems before the optimization. This is why we conclude, that chalcogen bond can be utilized in drug design to strengthen the binding to their targets.

   It has proven to be difficult to completely separate the contributions of chalcogen and halogen bonds.

8 References

Adhikari U, Scheiner S. (2014): Effects of charge and substituent on the S···N chalcogen bond. J Phys Chem A 118(17): 3183–3192.

Anilkumar GN, Lesburg CA, Selyutin O, Rosenblum SB, Zeng Q, Jiang Y, Chan TY, Pu H, Vaccaro H, Wang L, Bennett F, Chen KX, Duca J, Gavalas S, Huang Y, Pinto P, Sannigrahi M, Velazquez F, Venkatraman S, Vibulbhan B, Agrawal S, Butkiewicz N, Feld B, Ferrari E, He Z, Jiang CK, Palermo RE, McMonagle P, Huang HC, Shih NY, Njoroge G, Kozlowski JA. (2011): I. Novel HCV NS5B polymerase inhibitors: Discovery of indole 2-carboxylic acids with C3-heterocycles. Bioorg Med Chem Lett 21(18): 5336–5341.

Auffinger P, Hays FA, Westhof E, Ho PS. (2004): Halogen bonds in biological molecules. Proc Natl Acad Sci U S A 101(48): 16789–16794.

Bader RFW, Carroll MT, Cheeseman JR, Chang C. (1987): Properties of Atoms in Molecules: Atomic Volumes. J Am Chem Soc 109: 7968–7979.

Baumli S, Endicott JA, Johnson LN. (2010): Halogen bonds form the basis for selective P-TEFb inhibition by DRB. Chem Biol 17(9): 931–936.

Bauzá A, Quiñonero D, Deyàa PM, Frontera A. (2013): Halogen bonding versus chalcogen and pnictogen bonding: a combined Cambridge structural database and theoretical study. CrystEngComm 15: 3137-3144.

Bauzá A, Frontera A. (2015): Aerogen Bonding Interaction: A New Supramolecular Force? Angew Chem Int Ed Engl 54(25): 7340-7343.

Becker P, Cohen-Addad C, Delley B, Hirshfeld FL Lehmann MS. 1986. A Theoretical Study of Short S···O "Non-Bonded" Interactions. Smith VH. Jr. Henry F. Schaefer HF. Morokuma K (eds.). Applied Quantum Chemistry: 361-373. Springer Netherlands.

Benjahad A, Guillemont J, Andries K, Nguyen CH, Grierson DS. (2003): 3-iodo-4-phenoxypyridinones (IOPY's), a new family of highly potent non-nucleoside inhibitors of HIV-1 reverse transcriptase. Bioorg Med Chem Lett 13(24): 4309-4312.

Bent HA. Structural chemistry of donor-acceptor interactions. (1968): Chem. Rev 68(5): 587–648.

Bissantz C, Kuhn B, Stahl M. A medicinal chemist's guide to molecular interactions. (2010): J Med Chem 53(14): 5061-5084.

Bondi A. (1964): van der Waals Volumes and Radii. J Phys Chem 68(3): 441–451.

Bosshard HR, Marti DN, Jelesarov I. (2004): Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. J Mol Recognit 17(1): 1-16.

Bourne PE, Beran B, Bi C,Bluhm WF, Dimitropoulos D, Feng. Z, Goodsell DS, Prlic A, Quinn GB, Rose PW, Westbrook J, Yukich B, Young J, Zardecki C, Berman HM. (2011): RCSB Protein Data Bank Website, Evolution of. WIREs Comput Mol Sci 1: 782-789.

Brinck T, Murray JS, Politzer P. (1992): Surface electrostatic potentials of halogenated methanes as indicators of directional intermolecular interactions. Int J Quantum Chem 44: 57–64.

Brown CS, Lee MS, Leung DW, Wang T, Xu W, Luthra P, Anantpadma M, Shabman RS, Melito LM, MacMillan KS, Borek DM, Otwinowski Z, Ramanan P, Stubbs AJ, Peterson DS, Binning JM, Tonelli M, Olson MA, Davey RA, Ready JM, Basler CF, Amarasinghe GK. (2014): In silico derived small molecules bind the filovirus VP35 protein and inhibit its polymerase cofactor activity. J Mol Biol 426(10): 2045-2058.

Burling FT, Goldstein BM. (1992): Computational studies of nonbonded sulfur-oxygen and selenium-oxygen interactions in the thiazole and selenazole nucleosides J Am Chem Soc 114 (7): 2313–2320.

Carlsson ACC. Veiga AX, Erdélyi M. (2015): Halogen Bonding in Solution. Metrangolo P. Resnati G (eds.). Halogen Bonding II Volume Topics in Current Chemistry 359: 49-76. Springer International Publishing.

Carré F, Chuit C, Corriu RPG, Monforte P, Nayyar NK, Reyé C. (1995): Intramolecular coordination at phosphorus: donor-acceptor interaction in three- and four-coordinated phosphorus compounds. J Organomet Chem 499: 147–154.

Clark T, Hennemann M, Murray JS, Politzer P. J. (2007): Halogen bonding: the sigma-hole. Proceedings of "Modeling interactions in biomolecules II". Mol Model 13(2): 291-296.

Clark T, Murray JS, Lane P, Politzer P. (2008): Why are dimethyl sulfoxide and dimethyl sulfone such good solvents? J Mol Model 14(8): 689-697.

Cody V, Murray-Rust P. (1984): Iodine···X(O, N, S) intermolecular contacts: models of thyroid hormone protein binding interactions using information from the cambridge crystallographic data files. J Mol Struct 112: 189-199.

Coutard B, Decroly E, Li C, Sharff A, Lescar J, Bricogne G, Barral K. (2014): Assessment of Dengue virus helicase and methyltransferase as targets for fragment-based drug discovery. Antiviral Res 106: 61-70.

Cramer CJ, Gladfelter WL. (1997): Ab Initio Characterization of [H3N·BH3]2, [H3N·AlH3]2, and [H3N·GaH3]2. Inorg Chem 36(23): 5358–5362.

Custelcean R, Jackson JE. (2001): Dihydrogen bonding: structures, energetics, and dynamics. Chem Rev 101(7): 1963-1980

Ebrahimi A, Razmazma H, Delarami HS. (2016): The Nature of Halogen Bonds in [N···X···N]+ Complexes: A Theoretical Study. Phys Chem Res 1: 1-15.

Erdélyi M. (2012): Halogen bonding in solution. Chem Soc Rev 41(9): 3547-3557.

Fanfrlík J, Brynda J, Rezác J, Hobza P, Lepsík M. (2008): Interpretation of protein/ligand crystal structure using QM/MM calculations: case of HIV-1 protease/metallacarborane complex. J Phys Chem B 112(47): 15094-102.

Fang Y, Li AY, Ma FY. (2015): A comparative study of the chalcogen bond, halogen bond and hydrogen bond S···O/Cl/H formed between SHX and HOCl. J Mol Model 21(3): 61.

Fick RJ, Kroner GM, Nepal B, Magnani R, Horowitz S, Houtz RL, Scheiner S, Trievel RC. (2016): Sulfur-Oxygen Chalcogen Bonding Mediates AdoMet Recognition in the Lysine Methyltransferase SET7/9. ACS Chem Biol 11(3): 748-754.

Goldstein BM, Takusagawa F, Berman HW, Srivastava PC, Robins RK. (1985): Structural studies of a new antitumor and antiviral agent: selenazofurin and its alpha anomer. J Am Chem Soc 107 (5): 1394–1400.

Grabowski SJ. (2013a): Hydrogen and halogen bonds are ruled by the same mechanisms. Phys Chem Chem Phys 15(19): 7249-7259.

Grabowsi SJ. (2013b): Dihydrogen bond and X–H…σ interaction as sub-classes of hydrogen bond. J Phys Org Chem 26: 452–459.

Grasberger BL, Lu T, Schubert C, Parks DJ, Carver TE, Koblish HK, Cummings MD, LaFrance LV, Milkiewicz KL, Calvo RR, Maguire D, Lattanze J, Franks CF, Zhao S, Ramachandren K, Bylebyl GR,

Zhang M, Manthey CL, Petrella EC, Pantoliano MW, Deckman IC, Spurlino JC, Maroney AC, Tomczuk BE, Molloy CJ, Bone RF. (2005): Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. J Med Chem 48(4): 909-912.

Grimme S. (2011): Density Functional Theory with London Dispersion Corrections. WIREs Comput Mol Sci 1: 211-228.

Hardegger LA, Kuhn B, Spinnler B, Anselm L, Ecabert R, Stihle M, Gsell B, Thoma R, Diez J, Benz J, Plancher JM, Hartmann G, Isshiki Y, Morikami K, Shimma N, Haap W, Banner DW, Prof. Diederich F. Dr. (2011): Halogen bonding at the active sites of human cathepsin L and MEK1 kinase: efficient interactions in different environments. ChemMedChem 6(11): 2048-2054.

Harris PM, MackJr E, Blake FC. (1928): THE ATOMIC ARRANGEMENT IN THE CRYSTAL OF ORTHORHOMBIC IODINE. J Am Chem Soc 50(6): 1583–1600.

Hassel O. (1970): Structural aspects of interatomic charge-transfer bonding. Science: 170(3957): 497-502.

Himmel DM, Das K, Clark AD, Hughes SH, Benjahad A, Oumouch S, Guillemont J, Coupa S, Poncelet A, Csoka I, Meyer C, Andries K, Nguyen CH, Grierson DS, Arnold E. (2005): Crystal structures for HIV-1 reverse transcriptase in complexes with three pyridinone derivatives: a new class of nonnucleoside inhibitors effective against a broad range of drug-resistant strains. J Med Chem 48(24): 7582−7591.

Ho PS. (2014): Biomolecular halogen bonds. Top Curr Chem 358: 241-276.

Hobza P, Müller-Dethlefs K. (2010): Non-Covalent Interaction, Theory and Experiment. The Royal Society of Chemistry. Cambridge.

Howard EI, Sanishvili R, Cachau RE, Mitschler A, Chevrier B, Barth P, Lamour V, Van Zandt M, Sibley E, Bon C, Moras D, Schneider TR, Joachimiak A, Podjarny A. (2004): Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 A. Proteins 55(4): 792-804.

Huheey JE. (1965): The Electronegativity of Groups. J Phys Chem 69(10): 3284–3291.

Iwaoka M, Isozumi N. (2006): Possible roles of S···O and S···N interactions in the functions and evolution of phospholipase A2. BIOPHYSICS 2: 23-34.

Iwaoka M, Takemoto S, Tomoda S. (2002): Statistical and theoretical investigations on the directionality of nonbonded S...O interactions. Implications for molecular design and protein engineering. J Am Chem Soc 124(35): 10613-10620.

Jensen F. (2007): Introduction to Computational Chemistry, 2nd edition. John Wiley & Sons. Chichester.

Kodíček M, Karpenko V. (2013): Biofysikální chemie. Academia. Praha.

Kolář M, Hostaš J, Hobza P. (2014): The strength and directionality of a halogen bond are co-determined by the magnitude and size of the σ-hole. Phys Chem Chem Phys 16(21): 9987-9996.

Krogsgaard-Larsen P, Strømgaard K, Madsen U. (2010): Textbook of Drug Discovery and Design, 4th eidition. CRC press. Boca Raton.

Latimer WM, Rodebush WH. (1920): POLARITY AND IONIZATION FROM THE STANDPOINT OF THE LEWIS THEORY OF VALENCE. J.Am Chem Soc 42(7): 1419–1433.

Legon AC. (2010): The halogen bond: an interim perspective. Phys Chem Chem Phys 12(28): 7736-47.

Lemke CT, Goudreau N, Zhao S, Hucke O, Thibeault D, Llinàs-Brunet M, White PW. (2011): Combined X-ray, NMR, and kinetic analyses reveal uncommon binding characteristics of the hepatitis C virus NS3-NS4A protease inhibitor BI 201335. J Biol Chem 286(13): 11434-11443.

Lu Y, Li H, Zhu X, Zhu W, Liu H. (2011): How does halogen bonding behave in solution? A theoretical study using implicit solvation model. J Phys Chem A 115(17): 4467-4475.

Lu Y, Shi T, Wang Y, Yang H, Yan X, Luo X, Jiang H, Zhu W. (2009): Halogen bonding--a novel interaction for rational drug design? J Med Chem 52(9): 2854-2862.

Manna D, Mugesh G. (2012): Regioselective deiodination of thyroxine by iodothyronine deiodinase mimics: an unusual mechanistic pathway involving cooperative chalcogen and halogen bonding. J Am Chem Soc 134(9): 4269-4279.

Martinez CR, Iverson BL. (2012): Rethinking the term "pi-stacking" Chem Sci 3: 2191-2201.

Mitoraj MP, Michalak A. (2013): Theoretical description of halogen bonding - an insight based on the natural orbitals for chemical valence combined with the extended-transition-state method (ETS-NOCV). J Mol Model 19(11): 4681-4688.

Mitzel NW, Blake AJ, Rankin DWH. (1997): β-Donor Bonds in SiON Units: An Inherent Structure-Determining Property Leading to (4 + 4)-Coordination in Tetrakis-(N,N-dimethylhydroxylamido)silane. J. Am Chem Soc 119(18): 4143–4148.

Mó O, Lamsabhi AM, Yáñez M, Heverly-Coulson GS, Boyd RJ. (2013):.Dramatic substituent effects on the mechanisms of nucleophilic attack on Se-S bridges. J Comput Chem 34(29): 2537-2547.

Mulliken RS, Person WB. (1969): Molecular compounds and their spectra. XXI. Some general considerations. J Am Chem Soc 91(13): 3409–3413.

Murray JS, Lane P, Clark T, Politzer P. (2007): Sigma-hole bonding: molecules containing group VI atoms. J Mol Model 13(10): 1033-1038.

Murray JS, Lane P, Clark T, Riley KE, Politzer P. (2012): Σ-holes, π-holes and electrostatically-driven interactions. J Mol Model 18(2): 541-548.

Murray JS, Lane P, Politzer P. (2008a): Expansion of the sigma-hole concept. J Mol Model 15: 723-729.

Murray JS, Lane P, Politzer P. (2008b): Simultaneous σ-hole and hydrogen bonding by sulfur- and selenium-containing heterocycles. Int J Quantum Chem108: 2770–2781.

Murray JS, Politzer P. (2011): Electrostatic Potential: an Overview. WIREs Comput Mol Sci. 1: 153-163.

Murraya JS, Macaveiua L, Politzera P. (2014): Factors affecting the strengths of σ-hole electrostatic potentials J Comput Sci 5: 590–596.

Murray-Rust P, Motherwell WDS. (1979): Computer retrieval and analysis of molecular geometry. Intermolecular interactions. J Am Chem Soc 101(15): 4374–4376.

Palusiak M. (2010): On the nature of halogen bond – The Kohn–Sham molecular orbital approach. J Mol Struct 945: 89–92.

Parisini E, Metrangolo P, Pilati T, Resnati G, Terraneoab G. (2011): Halogen bonding in halocarbon–protein complexes: a structural survey. Chem Soc Rev 40: 2267-2278.

Parks DJ, Lafrance LV, Calvo RR, Milkiewicz KL, Gupta V, Lattanze J, Ramachandren K, Carver TE, Petrella EC, Cummings MD, Maguire D, Grasberger BL, Lu T. (2005): 1,4-Benzodiazepine-2,5-diones as small molecule antagonists of the HDM2-p53 interaction: discovery and SAR. Bioorg Med Chem Lett 15(3): 765-70.

Pavan MS, Jana AK, Natarajan S, Guru Row TN. (2015): Halogen Bonding and Chalcogen Bonding in 4,7-Dibromo-5,6-dinitro-2,1,3-benzothiadiazole. J Phys Chem B 119(34): 11382-11390.

Pecina A, Lepšík M, Řezáč J, Brynda J, Mader P, Řezáčová P, Hobza P, Fanfrlík J. (2013): QM/MM calculations reveal the different nature of the interaction of two carborane-based sulfamide inhibitors of human carbonic anhydrase II. .J Phys Chem B 117(50): 16096-16104.

Pendás MA, Blanco MA, Francisco E. (2006): The nature of the hydrogen bond: a synthesis from the interacting quantum atoms picture. J Chem Phys 125(18): 184112.

Politzer P, Murray JS, Clark T. (2010): Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. Phys Chem Chem Phys 12(28): 7748-5777.

Politzer P, Murray JS, Clark T. (2013): Halogen bonding and other σ-hole interactions: a perspective. Phys Chem Chem Phys 15(27): 11178-11189.

Politzer P, Murray JS, Concha MC. (2008): Sigma-hole bonding between like atoms; a fallacy of atomic charges. J Mol Model 14(8): 659-665.

Ramasubbu N, Parthasarathy R, Murray-Rust P. (1986): Angular preferences of intermolecular forces around halogen centers: preferred directions of approach of electrophiles and nucleophiles around carbon-halogen bond. J Am Chem Soc 108 (15): 4308–4314.

Ranjith-Kumar CT, Kao CC. (2006): Biochemical Activities of the HCV NS5B RNA-Dependent RNA Polymerase. Tan SL (ed.). Hepatitis C Viruses: Genomes and Molecular Biology. Horizon Bioscience 10. Norfolk.

Reed AE, Curtiss LA, Weinhold F. (1988): Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint Chem Rev 88(6): 899–926.

Richardson T, de Gala S, Crabtree RH, Siegbahn PEM. (1995): Unconventional Hydrogen Bonds: Intermolecular B-H...H-N Interactions. J Am Chem Soc 117(51): 12875–12876.

Riley KE, Murray JS, Fanfrlík J, Řezáč J, Solá RJ, Concha MC, Ramos FM, Politzer P. (2011): Halogen bond tunability I: the effects of aromatic fluorine substitution on the strengths of halogen-bonding interactions involving chlorine, bromine, and iodine. J Mol Model 17: 3309-3318.

Riley KE, Hobza P. (2008): Investigations into the Nature of Halogen Bonding Including Symmetry Adapted Perturbation Theory Analyses. J Chem Theory Comput 4(2):232-242.

Riley KE, Hobza P. (2013): The relative roles of electrostatics and dispersion in the stabilization of halogen bonds. Phys Chem Chem Phys 15(41):17742-17751.

Riley KE, Murray JS, Fanfrlík J, Rezáč J, Solá RJ, Concha MC, Ramos FM, Politzer P. (2013): Halogen bond tunability II: the varying roles of electrostatic and dispersion contributions to attraction in halogen bonds. J Mol Model 19(11): 4651-4659.

Robertson CC, Perutz RN, Brammer L, Hunter CA. (2014): A solvent-resistant halogen bond. Chem Sci 5: 4179-4183.

Sarwar MG, Dragisic B, Salsberg LJ, Gouliaras C, Taylor MS. (2010): Thermodynamics of halogen bonding in solution: substituent, structural, and solvent effects. J Am Chem Soc 132(5): 1646-1653.

Sedlak R, Fanfrlík J, Pecina A, Hnyk D, Hobza P, Lepšík M. (2015): Boron, The Fifth Element. Hnyk D, Michael L. McKee (eds.). Springer International Publishing. Cham, Heidelberg, New York, Dordrecht, London.

Shields ZP, Murray JS, Politzer P. (2010): Directional tendencies of halogen and hydrogen bonds. Int J Quantum Chem 110: 2823–2832.

Schimer J, Pávová M, Anders M, Pachl P, Šácha P, Cígler P, Weber J, Majer P, Řezáčová P, Kräusslich HG, Müller B, Konvalinka J. (2015): Triggering HIV polyprotein processing by light using rapid photodegradation of a tight-binding protease inhibitor. Nat Commun. 6: 6461.

Stone A. (2013): The Theory of Intermolecular Forces, OUP. Oxford.

Syzgantseva OA, Tognetti V, Joubert L. (2013): On the physical nature of halogen bonds: a QTAIM study. J Phys Chem A 117(36): 8969-9880.

Tsuzuki S, Sato N. (2013): Origin of attraction in chalgogen-nitrogen interaction of 1,2,5-chalcogenadiazole dimers. J Phys Chem B 117(22): 6849-6855.

Wang C, Danovich D, Mo Y, Shaik S. (2014): On The Nature of the Halogen Bond. J Chem Theory Comput 10(9): 3726-3737.

Wang C, Guan L, Danovich D, Shaik S, Mo Y. (2016): The origins of the directionality of noncovalent intermolecular interactions. J Comput Chem 37(1): 34-45.

Wang W, Hobza P. (2008): Origin of the X-Hal (Hal = Cl, Br) bond-length change in the halogen-bonded complexes. J Phys Chem A 112(17): 4114-4119.

Wang W, Ji B, Zhang Y. (2009): Chalcogen Bond: A Sister Noncovalent Bond to Halogen Bond J Phys Chem A 113 (28): 8132–8135.

Wang W, Wong NB, Zheng W, Tian A, (2004): Theoretical Study on the Blueshifting Halogen Bond. J Phys Chem A 108(10): 1799–1805.

Wilcken R, Zimmermann MO, Lange A, Zahn S, Boeckler FM. (2012): Using halogen bonds to address the protein backbone: a systematic evaluation. J Comput Aided Mol Des 26(8): 935-945.

Wilcken R, Zimmermann MO, Lange A, Joerger AC, Boeckler FM. (2013): Principles and applications of halogen bonding in medicinal chemistry and chemical biology. J Med Chem 56(4): 1363-1388.

Zachariasen WH, Mooney RCL. (1934): The Structure of the Hypophosphite Group as Determined from the Crystal Lattice of Ammonium Hypophosphite. J Chem Phys 2: 34.

Zhao Q. (2014): Interplay between halogen and chalcogen bonding in the XCl···OCS···NH_3 (X = F, OH, NC, CN, and FCC) complex. J Mol Model 20(10): 2458.

Zielinski Z, Presseau N, Amorati R, Valgimigli L, Pratt DA. (2014): Redox chemistry of selenenic acids and the insight it brings on transition state geometry in the reactions of peroxyl radicals. J Am Chem Soc 136(4): 1570-1578.

Zierkiewicz W, Bieńko DC, Michalska D, Zeegers-Huyskens T. (2015): On the nature of halogen bonded complexes between carbonyl bases and chlorotrifluoromethane. Theor Chem Acc 134: 103.

Zierkiewicz W, Hobza P. (2004): The dihydrogen bond in X3C–H···H–M complexes (X = F, Cl, Br; M = Li, Na, K). A correlated quantum chemical ab initio and density functional theory study. Phys Chem Chem Phys 6: 5288-5296.